Human Urinary Glycoproteomics; Attachment Site Specific Analysis of *N*- and *O*-Linked Glycosylations by CID and ECD*⑤

Adnan Halim‡, Jonas Nilsson‡, Ulla Rüetschi‡, Camilla Hesse‡, and Göran Larson‡§

**Urine is a complex mixture of proteins and waste products and a challenging biological fluid for biomarker discovery. Previous proteomic studies have identified more than 2800 urinary proteins but analyses aimed at unraveling glycan structures and glycosylation sites of urinary glycoproteins are lacking. Glycoproteomic characterization remains difficult because of the complexity of glycan structures found mainly on asparagine (*N*-linked) or serine/threonine (*O*-linked) residues. We have developed a glycoproteomic approach that combines efficient purification of urinary glycoproteins/glycopeptides with complementary MS-fragmentation techniques for glycopeptide analysis. Starting from clinical sample size, we eliminated interfering urinary compounds by dialysis and concentrated the purified urinary proteins by lyophilization. Sialylated urinary glycoproteins were conjugated to a solid support by hydrazide chemistry and trypsin digested. Desialylated glycopeptides, released through mild acid hydrolysis, were characterized by tandem MS experiments utilizing collision induced dissociation (CID) and electron capture dissociation fragmentation techniques. In CID-MS^2, Hex$_5$HexNAc$_4$-*N*-Asn and HexHexNAc-*O*-Ser/Thr were typically observed, in agreement with known *N*-linked biantennary complex-type and *O*-linked core 1-like structures, respectively. Additional glycoforms for specific *N*- and *O*-linked glycopeptides were also identified, *e.g.* tetra-antennary *N*-glycans and fucosylated core 2-like *O*-glycans. Subsequent CID-MS^3, of selected fragment-ions from the CID-MS^2 analysis, generated peptide specific *b*- and *y*-ions that were used for peptide identification. In total, 58 *N*- and 63 *O*-linked glycopeptides from 53 glycoproteins were characterized with respect to glycan- and peptide sequences. The combination of CID and electron capture dissociation techniques allowed for the exact identification of Ser/Thr attachment site(s) for 40 of 57 putative *O*-glycosylation sites. We defined 29 *O*-glycosylation sites which have, to our knowledge, not been previously reported. This is the first study of human urinary glycoproteins where "intact" glycopeptides were studied, *i.e.* the presence of glycans and their attachment sites were proven without doubt.   *Molecular & Cellular Proteomics 11: 10.1074/mcp.M111.013649, 1–17, 2012.*

In search of disease biomarkers, urine qualifies as an important biologic fluid that can easily be collected by repeated and noninvasive sampling from single individuals. Proteins present in urine are derived not only from glomerular ultrafiltration of plasma but also from tubular secretion of soluble proteins, detachment of glycosylphosphatidyl inositol anchored proteins and exosome shedding through the urothelium (1). For healthy individuals, 30% of the urinary proteome has been estimated to originate from the plasma filtrate whereas the remaining 70% is believed to be derived from the kidneys and the urothelium (2). Until 2005, ∼800 urinary proteins had been identified by various proteomic approaches (3–7). In 2006, a comprehensive proteomic study identified more than 1500 proteins from healthy human urine samples, simultaneously reflecting the complexity and the potential information concealed in the urinary proteome (8). In 2009, Kentsis *et al.* reported the hitherto largest data set for the urinary proteome, unveiling more than 2300 protein identities (9). The "core urinary proteome" was recently defined as a common set of nearly 600 urinary proteins with a dynamic concentration range spanning five orders of magnitude (10). Interestingly, the authors also reported that the 20 most abundant proteins, which were estimated to constitute 2/3 of the core urinary proteome by mass, were glycoproteins with serum albumin being the only exception.

Glycoproteins are characterized by the presence of oligosaccharides linked to the peptide backbone primarily through *N*- or *O*-glycosidic bonds at asparagine or serine/threonine residues, respectively (11). *N*- and mucin-type *O*-glycosylations are widely accepted as the most common and structurally diverse post-translational modifications found on secreted proteins and on the extracellular parts of membrane bound proteins (12). Given that protein glycosylation is involved in various cellular processes (13–16), the site-specific characterization of *N*- and *O*-linked glycosylations and identification of the modified proteins is becoming increasingly important. Urine is potentially a rich source for *N*- and *O*-linked glycoproteins derived from renal- and distal organs and represents an interesting subproteome for structural charac-

terization of human glycoproteins. However, glycoproteomic characterization of urine is lacking and only a few proteomic studies aimed at identifying urinary glycoproteins have been reported (17–20). In these studies, the glycan moieties were either cleaved off or not studied at all. It is, however, important to analyze qualitative glycan differences in glycoproteomes because changes associated with the carbohydrate moieties may reflect physiological status (21–23). Perhaps more importantly for the urinary proteome, the study of intact glycopeptides could reveal not only the glycoprotein origin but potentially also provide information regarding pathological changes of its original tissue (24, 25). By analyzing tryptic glycopeptides originating from urinary glycoproteins both the glycan structures and glycosylation sites of proteins may be addressed. However, a highly purified mixture of glycopeptides is the prerequisite for such studies because of the general phenomena of ion suppression and stoichiometric effects in the mass spectrometric analysis of complex mixtures (26–28). Enrichment methods for the isolation of formerly *N*-linked glycopeptides from biological sources have been described using hydrazide chemistry, TiO_2 affinity purification, lectin chromatography and hydrophilic interaction liquid chromatography (HILIC) (29–33). The *N*-glycans are typically removed by PNGase F treatment during these protocols and the site-specific information of *N*-glycan structures is usually not addressed. Only a few glycoproteomic studies, aimed at analyzing intact *N*-glycopeptides from biological samples, have been published (34, 35). Also, by comparison to *N*-glycosylation, characterization of protein *O*-glycosylation is analytically more challenging for several reasons, *e.g.* due to the heterogeneity associated with *O*-glycan core structures (36). Although collision-induced dissociation (CID)[1]-based MS^n strategies are well capable of revealing both *O*-glycan- and peptide sequences for intact glycopeptides (37) the site-specific information of the modified amino acid is however usually lost. This is because of predominant glycosidic fragmentation of the precursor during MS^2, and peptide fragmentation occurring mainly for the deglycosylated peptide ion in the MS^3. Additionally, the exact glycosylation site of identified peptides containing several Ser/Thr residues cannot be predicted due to the lack of a consensus sequence for mucin-type *O*-glycosylation. The alternative fragmentation techniques electron capture dissociation (ECD) (38, 39) and electron transfer dissociation (ETD) (40) have been introduced for site-specific analysis of CID-labile PTMs but characterization of protein *O*-glycosylations using ECD/ETD have generally been limited to synthetic glycopeptides or single glycoproteins (41–45).Thus, investigation of protein *O*-glycosylation has lagged

behind and relatively little is known about *O*-linked glycans with respect to their protein carriers and amino acid attachment sites. Recently, Darula and Medzihradszky used lectin enrichment with jacalin, recognizing core 1 *O*-glycans (Galβ1–3GalNAcα-*O*-Ser/Thr), and identified 21 *O*-glycosylation sites from bovine serum glycoproteins by combining ETD and exoglycosidase digestion (46). We have previously developed a sialic acid specific capture-and-release protocol for the enrichment of both *N*- and *O*-glycosylated peptides from sialylated glycoproteins in biological samples using hydrazide chemistry (37). Only CID based characterization was employed in our previous study and assignment of *O*-glycan attachment sites was therefore not possible for most *O*-glycosylated peptides. The low sensitivity and fragmentation yield for ECD/ETD compared with CID make it advantageous to use highly enriched samples of *O*-glycosylated peptides. We tested the sialic acid capture-and-release protocol on human serum samples but, as expected, *N*-glycosylated peptides completely dominated the LC-MS/MS chromatograms (Halim *et al.*, unpublished). We then turned our attention to urine, with ambitions to characterize *N*- and *O*-glycosylated peptides, since urine also may serve as a sample source for biomedical diagnosis. However, because urine contains much salts and pigments, which could interfere with the periodate oxidation step in our protocol, we first developed a simple method to remove low-molecular waste products and attain pure protein samples suitable for redox chemistry and proteomics purposes. In this study, we have thus extended our protocol (Fig. 1 and supplemental Fig. S1) to include a unique dialysis procedure for isolation of human urinary proteins prior to the sialic acid capture-and-release method. In addition to the CID-based approach, we also included ECD for the characterization of *O*-glycan attachment sites and as a complementary peptide fragmentation mode for the identification of urinary glycopeptides.

EXPERIMENTAL PROCEDURES

Collection and Preparation of Human Urine—First morning, midstream urine was obtained from a healthy male individual during five consecutive days and prepared separately. Immediately after collection, 50 ml de-identified urine was separated from intact cells and debris by centrifugation at 3000 $\times$ *g*, 4 °C for 20 min. The uppermost 20 ml were frozen at −20 °C and used for further analysis. Routine clinical chemistry analyses of all five samples were all within the reference range (U-Albumin (<5,4 mg/L), U-Creatinine (mean 17,4 mmol/L; range 12–28 mmol/L). U-Bilirubin, U-Urobilinogen, U-Acetone, U-Glucose, U-Erythrocytes, U-Leukocytes, U-nitrite were all negative).

After thawing, 10 ml of each sample was dialyzed against 14 $\times$ 2 L of tap water at 4 °C using Spectra/Por MWCO 12–14 kDa (Spectrum Laboratories) for 7 days (Fig. 1). The urine samples were lyophilized, dissolved in 6 ml 5% sodium-dodecyl sulfate (SDS) and dialyzed against 2 $\times$ 2 L of 1.5% SDS at 60 °C for 24 h. The SDS was subsequently removed by dialysis against 2 $\times$ 2 L Milli Q deionized H_2O (dH_2O) at room temperature for 24 h. Finally, the samples were lyophilized and dissolved in 0.5 ml dH_2O. Protein content was determined using the BCA-1 protein assay (Sigma-Aldrich) on a NanoDrop

---

[1] The abbreviations used are: Con A, concanavalin A; dHex, deoxyhexose; ECD, electron capture dissociation; Gal, galactose; GalNAc, N-acetylgalactosamine; GlcNAc, N-acetylglucosamine; Hex, hexose; HexNAc, N-acetylhexosamine; PNGase F, peptide N-glycosidase F; WGA, wheat germ agglutinin.

1000 spectrophotometer (Thermo Scientific) according to the manufacturer's protocol.

Protein Separation—For protein separation prior to in-gel trypsin digestion 80 μg of urinary proteins were dissolved in NuPage LDS-sample buffer (Invitrogen, Carlsbad, CA) supplemented with 50 mM dithiothreitol, reduced and denatured at 70 °C for 10 min. Protein samples were then separated on 4–12% Bis-Tris precasted polyacrylamide gels (Invitrogen). SeeBlue Plus2 pre-stained standard (Invitrogen) was used as molecular weight marker and proteins were visualized by Coomassie colloidal blue staining. For in-gel trypsin digestion one gel lane was divided into 15 equally sized gel slices and subjected to automated trypsin digestion (supplemental Fig. S1A) on a BioMek 2000 work station equipped with a vacuum manifold. 96-well plates supplemented with a 7 μl volume of C18 reversed phase chromatographic resin were used for vacuum filtration and sample clean-up. The work-flow essentially followed the protocol previously described (47) except that the peptide extraction was performed twice with 0.2% trifluoroacetic acid to allow for peptide binding to the C18 resin of the filter plates. Finally, peptides were eluted twice in 40 μl of 60% acetonitrile in 0.1% trifluoroacetic acid and the eluted fractions were evaporated to dryness in a vacuum centrifuge. Prior to liquid chromatography/tandem MS (LC-MS/MS) analysis samples were redissolved in 0.1% formic acid.

For electrophoretic analysis of repeatedly dialyzed urine samples, 30 μg of urinary proteins were denatured by heating (100 °C, 5 min) in 1% SDS and 100 mM dithiothreitol and separated on a 4–12% Bis-Tris precasted polyacrylamide gel (Invitrogen). SeeBlue Plus2 prestained standard (Invitrogen) was used as molecular weight marker and proteins were visualized by Coomassie colloidal blue staining (supplemental Fig. S2C).

Glycopeptide Enrichment Procedure

Hydrazide Capture—Capture of sialylated glycoproteins to hydrazide beads (supplemental Fig. S1B) was done as previously described (37) with minor modifications. One hundred μg protein in 1 ml dH$_2$0 was oxidized with 2 mM periodic acid for 15 min at 0 °C. The reaction was quenched by the addition of 5 μl 99% glycerol and buffer exchanged to 2.5 ml coupling buffer (100 mM acetate, 150 mM NaCl, pH 4.5) using Sephadex PD-10 columns (GE Healthcare). One hundred μl hydrazide beads (Bio-Rad) in coupling buffer was added and agitated for 16 h at room temperature in the dark. The beads were subsequently washed with 3 × 3 ml 0.1% Tween 20 in PBS, pH 7.4 and finally with 2 × 3 ml of 50 mM NH$_4$HCO$_3$, pH 8.0.

Reduction, Alkylation and Trypsin Digestion—The glycoproteins captured onto the beads were then incubated with 0.3 ml 10 mM dithiothreitol for 1 h at 37 °C in the dark. Following a washing step (50 mM NH$_4$HCO$_3$, pH 8.0), 0.3 ml 55 mM iodoacetamide (Sigma Aldrich) was added and incubated for 30 min at room temperature and in the dark. The beads were then washed with 2 × 3 ml of 8 M urea, 50 mM NH$_4$HCO$_3$, pH 8.0 and with 2 × 3 ml of 1% SDS in dH$_2$0 with gentle agitation. Finally, five washing steps with 3 ml of 50 mM NH$_4$HCO$_3$, pH 8.0 were performed. Captured glycoproteins were digested with 1 μg sequencing grade porcine trypsin (Promega, Madison, WI) in 70 μl 50 mM NH$_4$HCO$_3$, pH 8.0, at 37 °C for 18 h. The trypsin-released peptides were transferred to prelubricated eppendorf tubes (Costar). Any remaining peptides were extracted once with 100 μl 50% acetonitrile, pooled and lyophilized together with the trypsin released peptides and subjected to mass spectrometric analysis (supplemental Fig. S1B).

Release of Glycopeptides—The beads were initially washed once with 3 ml of 50% acetonitrile in dH$_2$0, once with 3 ml dH$_2$0 and once with 3 ml 1.5 M NaCl in dH$_2$0. The beads were then washed 3 × 3 ml dH$_2$0, 2 × 3 ml 50% acetonitrile in dH$_2$0, 2 × 3 ml with 25% acetonitrile in dH$_2$0 and finally with 2 × 3 ml dH$_2$0. One hundred μl 0.1 M formic acid was added to the beads and incubated for 1 h at 80 °C

(supplemental Fig. S1C). The released glycopeptides were transferred to prelubricated eppendorf tubes (Costar, Cambridge, MA). Any remaining glycopeptides were extracted once with 50 μl 50% acetonitrile in dH$_2$0, pooled and lyophilized together with the formic acid released glycopeptides and subjected to mass spectrometric analysis.

LC-MS/MS Analysis—Tryptic peptides, obtained either from in-gel digestion of electrophoretically separated urinary proteins (supplemental Fig. S1A), from unglycosylated peptides released by trypsin digestion of hydrazide captured glycoproteins (supplemental Fig. S1B) or glycopeptides released through formic acid hydrolysis (supplemental Fig. S1C) were separated by reversed phase chromatography on a 15 cm capillary column (Zorbax SB300 C18, 0.075 mm ID). Peptides/glycopeptides were reconstituted in 40 μl 0.1% formic acid, 20 μl was loaded onto the column in eluent A (0.1% formic acid) and separated with a linear gradient from 3% to 60% eluent B (84% acetonitrile in 0.1% formic acid) at a flow rate of 250–300 nL/min. Gradient lengths were either 50 min, for the analysis of the peptide fraction, or 150 min, for the glycopeptide fraction and the in-gel digested fractions. The LC system (Ettan MDLC, GE Healthcare) was coupled in-line with a LTQ-FTICR instrument (Thermo Fisher Scientific) via a nanoelectrospray source (Thermo Fisher Scientific). The source was operated at 1.4 kV, with no sheath gas flow and with the ion transfer tube at 200 °C. The mass spectrometer was programmed for acquisition in a data dependent mode. The survey scans were acquired in the FTICR mass analyzer and covered the *m/z* range 300–2000. For the analysis of peptides the seven most intense peaks in each full mass scan, with charge state $\geq$2 and intensity above a threshold of 100, were selected for fragmentation in the linear ion trap (LTQ) by CID. Glycopeptides were analyzed with two independent methods, one based on CID fragmentation and the other on ECD fragmentation. For the CID method the most intense peak in each FTICR full scan was selected for fragmentation in the linear ion trap (LTQ) followed by subsequent selection and fragmentation of the five most intense MS^2 fragment ions. For the ECD method the two most intense peaks in each FTICR full scan was selected for fragmentation in the ICR cell. CID fragmentation was performed with normalized collision energy of 35% activation, q = 0.25, activation time of 30 ms and three microscans. ECD fragmentation was performed with a relative energy of 4 and 5 in subsequent scans and a duration of 70 ms and three microscans. For all fragmentation events dynamic exclusion was enabled with a repeat count of 2. Peaks selected for fragmentation more than twice within a 30 s interval were excluded from selection (20 ppm window) for 180 s and the maximum number of excluded peaks was 200. AGC settings were 1000000 (FTMS full scan), 30000 (Ion trap), 10000 (Ion trap MS^n), and 500000 (FTMS ECD).

Data Analysis

Protein Identification—Raw data containing centroid MS/MS spectra, from the analysis of tryptic peptides, were converted into .dta format by the Bioworks software (version 3.3.1) utility extract_msn (Thermo Fisher Scientific) and analyzed with an in-house version of the Mascot software (Mascot ver. 2.3.01, http://www.matrixscience.com). Search parameters were set as follows: peptide tolerance, 10 ppm; MS/MS tolerance, 0.5 Da; enzyme, trypsin, one missed cleavage allowed; fixed carbamidomethyl modification of cysteine; variable oxidation of methionine; database, IPI human version 3.72 (86,392 sequences). Fragment ions from the *b*- and *y*-series, including losses of ammonia or water, were used for scoring. Minimal requirement for each protein identification was two unique peptide hits with scores above the significance threshold ($p < 0.05$).

Protein Clustering—Mascot results, including information on identified proteins and peptides, were imported into the ProteinCenter

software (Proxeon Bioinformatics). Data was filtered so that each identified protein contained at least two unique peptides and identified proteins were clustered, based on peptide sharing, into groups of indistinguishable proteins. Lists of protein identifiers from two independent studies (8, 9) were also imported into the ProteinCenter software and comparisons of the three data sets were performed.

Glycopeptide Characterization Using CID—Glycopeptide identification and relative quantification of *N*- and *O*-glycan microheterogeneity was done as previously described (37). *N*- and *O*-linked glycan sequences were manually verified in CID-MS^n spectra for each glycopeptide by tracing peaks corresponding to the loss of individual monosaccharides. Manually selected MS^3 spectra, corresponding to the fragmentation of unmodified peptides for *O*-glycopeptides, were individually converted to .mzXML format via the Readw application (http://www.proteomecenter.org). Each .mzXML file was individually visualized with the mMass (version 2.4) application (48) and searched with the Mascot algorithm. The peptide monoisotopic mass was manually defined for each search by subtracting the monoisotopic mass of the glycan from the FTICR-MS^1 measured precursor. Search parameters were set as follows: peptide tolerance, 10 ppm; MS/MS tolerance, 0.6 Da; enzyme, trypsin, one missed cleavage allowed; fixed carbamidomethyl modification of cysteine; variable oxidation of methionine and variable loss of NH$_3$ (-17.0266 Da) at N-terminal cysteine and glutamine; taxonomy, human, 20,259 sequences (protein entries); database, SwissProt 101005. Peptides were considered as positive identifications if the ion score was above the significance threshold ($p < 0.05$). For MS^3 spectra that did not yield positive identifications, in the above described procedure, the peak list of individual glycopeptides were manually exported from the mMass application as .txt files and analyzed with an in-house version of the Mascot software (Mascot version 2.3.01, www.matrixscience.com). The precursor mass was manually defined in each .txt file so that it would match the monoisotopic mass of the peptide as described above. Enzyme specificity was set to semitrypsin or to no enzyme to account for peptides with a single or no tryptic sites, respectively. Finally, variable phosphorylation at serine or threonine residues was used in selected cases. All CID-MS^3 spectra that resulted in positive identifications were also converted to .mgf files according to the same procedures as above and Mascot searched against a decoy database (taxonomy, human, 20,245 sequences (protein entries); database, Swissprot 110817) using the same search parameters as above.

For all *N*-linked glycopeptides, the peak list for CID-MS^3 spectra of selected ions (peptide+HexNAc or peptide+dHexHexNAc) was converted to .txt files as described above. The precursor mass was manually defined in each .txt file so that it would match the monoisotopic mass of the peptide+HexNAc or peptide+dHexHexNAc. This was accomplished by subtracting the monoisotopic mass of the *N*-glycan (apart from HexNAc or dHexHexNAc) from the monoisotopic mass of the FTICR-MS^1 measured precursor. The sequence rule SEQ = B-NX[STC] or SEQ = C-N[KR] was included in the .txt file to constrain each search against peptide sequences containing the *N*-glycosylation consensus (with or without a tryptic cleavage site within the consensus sequence itself). This constraint lowered the acceptance threshold value but was justified by the clear presence of the *N*-linked glycan sequence in CID-MS^2. Search parameters were as described above, with the exception of including HexNAc (203.0794 Da) or dHexHexNAc (349.1373 Da) as variable modification of asparagine. Mascot scoring options were set to include the neutral loss of HexNAc (203.0794 Da) from the precursor ion and from peptide b- and y-type fragments. Searches were performed with the Mascot algorithm and peptides were considered as positive identifications if the ion score was above the significance threshold ($p < 0.05$). All CID-MS^3 spectra that resulted in positive identifications

were also converted to .mgf files according to the same procedures as above and Mascot searched against a decoy database (taxonomy, human, 20,245 sequences (protein entries); database, Swissprot 110817) using the same search parameters as described for *N*-linked glycopeptides above.

Glycopeptide Characterization Using ECD—The precursor ion masses of ECD spectra were matched to precursor ion masses of glycopeptides that had been identified by the CID-MS^n approach. Peak lists of *c*, (*c* - 1), *z* and (*z*+1)-ions were prepared for candidate glycopeptides using the MS-product tool (http://prospector.ucsf.edu). Glycopeptide identifications were verified and *O*-glycan attachment sites were pinpointed manually to unique Ser/Thr residues by tracing *c*- and *z*-ion peaks that contained or lacked the anticipated glycan(s). Also, the Mascot distiller program (version 2.3.2.0, Matrix Science) was used for peak picking and to prepare Mascot files from the ECD spectra. Subsequent MS^2 spectra at relative energy 4 and 5 were aggregated and the ions presented as singly protonated in the output Mascot files. Search parameters were set as follows: peptide tolerance, 10 ppm; MS/MS tolerance, 0.03 Da; enzyme, trypsin, one missed cleavage allowed; fixed carbamidomethyl modification of cysteine; variable modification of HexHexNAc (365.1322 Da), Hex$_2$HexNAc (730.2644 Da) and dHexHex$_2$HexNAc$_2$ (876.3223 Da) of serine, threonine and tyrosine; variable Hex$_5$HexNAc$_4$ (1622.5816 Da) modification of asparagine; variable oxidation of methionine and variable loss of NH$_3$ (-17.0266 Da) at N-terminal cysteine and glutamine; taxonomy, human (20, 259 sequences); database, SwissProt 101005. *Instrument* was set to match 1+ ions of the *c*, *z* and *z*+1 series (*c*, *z*+1 and *z*+2 using Mascot terminology). We did not observe any y-ions and these were thus not considered in the scoring. Acceptance criteria for a positive identification was based on scoring above the significance threshold value ($p < 0.05$). The Mascot files were analyzed with the in-house version of the Mascot software (Mascot version 2.3.01).

RESULTS

Protein Yields and Identifications—Starting from 10 ml urine we used dialysis against water to remove salts and pigments but this was found to yield inadequate sample purity. However, after a second dialysis against 1.5% SDS at 60 °C the procedure was satisfactorily efficient in removing pigments (Fig. 1 and supplemental Fig. S2). We recovered 31 ± 10 μg/ml protein (mean ± 1SD) from the dialyzed urine samples. One dialyzed urine sample was analyzed by GeLC-MS/MS (supplemental Fig. S1A). Applying the criteria of at least two uniquely identified peptides per identified protein, we identified 989 urinary proteins that were grouped into 413 protein groups of indistinguishable proteins by clustering based on peptide sharing (Supplementary excel Table, Gel-based proteomics). Following hydrazide capture (supplemental Fig. S1B), 63 proteins were either identified only from peptides found in the tryptic digests of captured proteins ($n = 10$), only from the covalently linked glycopeptides released through acid hydrolysis ($n = 36$) (supplemental Fig. S1C) or from both of these procedures ($n = 17$). Thus, 53 glycoproteins could be identified solely based on the identification of unique glycopeptides and for 17 of those glycoproteins the identities were also supported by peptide identifications (supplemental Table S1 and supplemental Fig. S3). Altogether, 26 urinary glycoproteins were identified from 122 un-
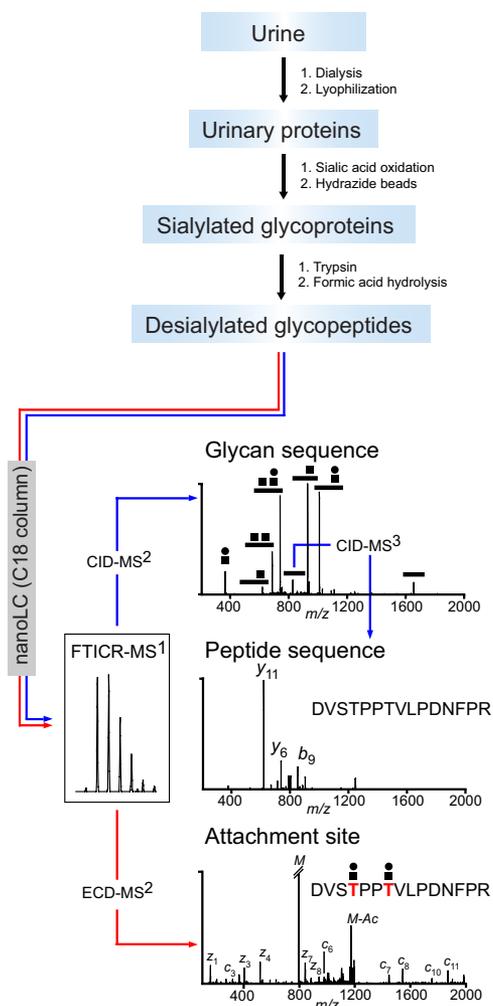
FIG. 1. **Schematic workflow for preparation of urinary proteins, enrichment of sialylated glycoproteins, release of desialylated glycopeptides and their mass spectrometric characterization by CID and ECD.**

glycosylated peptides found in the tryptic digests of glycoproteins captured onto the beads. Most of these proteins were annotated either as glycoproteins ($n = 20$) or as potential glycoproteins ($n = 4$) in the UniProtKB/Swiss-Prot database (49), *e.g.* Uromodulin, Kallikrein-1, Kininogen, Zinc-alpha-2-glycoprotein etc. Also, Phosphoinositide-3-kinase-interacting protein 1 (UniProt/KB accession Q96FE7) and Protein YIPF3 (UniProt/KB accession Q9GZM5), which are currently not annotated as potential glycoproteins, were indeed found to be glycosylated (see below). Serum albumin repeatedly appeared together with the enriched glycoproteins and was identified from 18 peptides only in the tryptic digests of the beads. In total, 442 urinary protein groups were identified in our samples by gel-based proteomics and hydrazide capture enrichment. We observed 400 protein identifications overlapping the data sets of Kentsis *et al.* and Adachi *et al.*, whereas 42 protein identifications were found to be unique in our data set (supplemental Fig. S3).

Identification of O-Linked Glycopeptides by CID—We identified 63 glycopeptides, corresponding to 49 differently *O*-glycosylated peptides originating from 40 urinary glycoproteins (0.0% false positive identifications). These are presented, together with their identified *O*-linked glycans, their attachment sites and Mascot scores of individual glycopeptides in Table I. Annotated CID-MS^n and ECD spectra for each *O*-glycopeptide is presented in supplemental Fig. S5. The relative abundance of specific glycoforms at each *O*-glycan attachment site are listed in supplemental Table S2. Typical CID-MS^n experiments for three *O*-linked glycopeptides constituting the same tryptic peptide are presented in Figs. 2*A*–2*C* to illustrate our strategies for glycan fragmentation analysis and manual identification of *O*-linked glycopeptides. Doubly (dashed line) and triply (solid line) protonated precursor ions of co-eluting glycoforms in the ion chromatograms (Fig. 2*E* and supplemental Fig. S4) were accurately mass measured (± 10 ppm) in the ICR cell. CID-MS^2 of the Hex$_2$HexNAc$_2$ glycoform resulted in fragmentation of the glycan part into Y-type ions (nomenclature according to Domon and Costello (50) and a B-type ion corresponding to the [HexHexNAc$+$H]$^+$ oxonium ion at *m/z* 366 (Fig. 2A). The most abundant Y-type ions were frequently observed as charge reduced species, *e.g.* at *m/z* 1010.4 and *m/z* 929.8 (Fig. 2A and supplemental Fig. S5). The neutral loss of one and two Hex residues (*m/z* 741.7 and *m/z* 687.8, respectively) followed by the loss of one HexNAc residue (*m/z* 620.3 and *m/z* 929.8) and finally the loss of the final HexNAc residue (*m/z* 828.4) demonstrated the Hex$_2$HexNAc$_2$ composition. For Hex$_2$HexNAc$_2$ containing glycopeptides we could not distinguish two separate HexHexNAc-*O*-Ser/Thr core 1-like glycans from one Hex(HexHexNAc)HexNAc-*O*-Ser/Thr core 2-like structure solely based on the Y-type ions. In CID-MS^2 and MS^3, these glycoforms could be differentiated by the presence of glycan fragments (B and internal B/Y-type ions) exceeding the HexHexNAc (*m/z* 366) composition, *e.g.* by the presence of diagnostic ions at *m/z* 407 corresponding to an internal Hex-NAcHexNAc fragment (51). The CID-MS^2 spectrum of the Hex$_2$HexNAc$_2$ glycoform in Fig. 2A did not contain a fragment ion at *m/z* 407, suggesting that two separate core 1-like glycans occupied two individual Ser/Thr residues within the glycopeptide. Conversely, in other cases core 2-like glycans were indeed identified (Fig. 3, see below). For the Hex$_2$HexNAc$_2$ glycoform (Fig. 2*A*) the intact peptide ion (Y$_0$-ion) was observed as the fifth most intense ion (for $z \geq 2$ ions) at *m/z* 828.4 and peptide fragmentation was obtained in the final CID-MS^3 spectrum. The HexHexNAc$_2$ glycoform was the next glycopeptide that eluted (*m/z* 741.7, Fig. 2E) and the CID-MS^2 spectrum (Fig. 2B) showed an intense charge reduced fragment ion at *m/z* 929.1 corresponding to the loss of HexHexNAc and a proton from the precursor ion. Additional charge reduced fragment ions at *m/z* 1010.6 and 827.7 showed the loss of HexNAc and HexHexNAc$_2$, respectively. CID-MS^3 of Y$_0$ at *m/z* 827.7 resulted in peptide fragmentation

TABLE I

List of urinary glycoproteins identified from their glycopeptide sequences, glycan structures and exact attachment sites.

*Bold and underlined residue depicts glycan attachment site, underlined residues depict experimentally indistinguishable attachment sites. * p < 0.05.*

Uniprot/KB accession	Glycoprotein	Peptide sequence	Attachment site	Glycan	Mascot score	Mascot threshold*	ECD
O-linked Glycopeptides							
P02765	Alpha-2-HS-glycoprotein	R.TVVQP**S**VGAAAGPVVPPCPGR.I	346	HexHexNAc	25	>18	Yes
P02656	Apolipoprotein C-III	D.PEVRP**T**SAVAA.-	94	HexHexNAc	19	>16	Yes
Q13790	Apolipoprotein F	K.DANISQPET**T**KEGLR.A	256^b	HexHexNAc	57	>21	Yes
P98160	Basement membrane-specific heparan sulfate proteoglycan core protein	R.AYDGLSLPEDIETV**T**ASQMR.W	42^b	HexHexNAc	93	>20	Yes
P26842	CD27 antigen	D.PLPNP**S**LTAR.S	127^b	HexHexNAc	16	>14	Yes
P16070	CD44 antigen	S.QEGGAN**T**TSGPIR.T	637–638^b	HexHexNAc	43	>33	Yes
P00742	Coagulation factor X	R.SVAQATSSSGEAPDS**I**TWKPYDAADLD.P	183–203	HexHexNAc	41	>38	No
P00742	Coagulation factor X	K.**S**HAPEVITSSPL.K	476–485^b	HexHexNAc	37	>33	No
P39059	Collagen alpha-1(XV) chain	E.ILEAVTY**T**QASPK.E	265^b	HexHexNAc	68	>32	Yes
P10643	Complement component C7	N.PL**T**QAVPK.C	696^b	HexHexNAc	39	>29	No
Q68CJ9	Cyclic AMP-responsive element-binding protein 3-like protein 3	R.VAADAVPG**S**EAPGPRPEADTTR.E	379^b	HexHexNAc	47	>21	Yes
Q13508	Ecto-ADP-ribosyltransferase 3	K.SQGNINNP**T**PGPVPVPGPK.S	346^b	HexHexNAc	59	>20	Yes
P98095	Fibulin-2	R.AEAGARPEENLILDAQA**T**SR.S	347–348^b	HexHexNAc	54	>24	No
Q8NFU4	Follicular dendritic cell secreted peptide	R.RNFPIPIPESAPT**T**PLP**S**E.K	75–83^b	HexHexNAc	64	>31	No
P78423	Fractalkine	K.AQDGGPVG**T**ELFR.V	183^b	HexHexNAc	79	>22	Yes
P78423	Fractalkine	R.VWGQGQ**S**PRPENSLER.E	253	HexHexNAc	35	>21	Yes
P78423	Fractalkine	R.LGVLITPVPDAQAATR.R	329–338	HexHexNAc	43	>13	No
P04921	Glycophorin-C	D.PGM**S**GWPDGR.M	42^a	HexHexNAc	59	>30	No
Q8TDQ0	Hepatitis A virus cellular receptor 2	R.DF**T**AAFPR.M	145^a	HexHexNAc	32	>21	No
P04233	HLA class II histocompatibility antigen gamma chain	S.LEQKP**T**DAPPK.V	203^b	HexHexNAc	37	>32	No
P04233	HLA class II histocompatibility antigen gamma chain	D.PSSGLGVTKQDLGPVP.M.-	281–287^b	HexHexNAc	42	>40	No
P01344	Insulin-like growth factor II	R.DVS**T**PP**T**VLPDNFPR.Y	96^b, 99^a	$Hex_2HexNAc_2$	67	>21	Yes
P01344	Insulin-like growth factor II	R.DVS**T**PP**T**VLPDNFPR.Y	99^a	HexHexNAc	63	>21	Yes
P01344	Insulin-like growth factor II	R.DVS**T**PP**T**VLPDNFPR.Y	96^b–99^a	$HexHexNAc_2$	39	>21	Yes
P01344	Insulin-like growth factor II	P.LIALP**T**QD.P	163	HexHexNAc	(38)	>42	No
P19823	Inter-alpha-trypsin inhibitor heavy chain H2	K.VVPDSTP**S**WANPSP**T**PVI**S**ML.A	665–679^b	$Hex_3HexNAc_3$	(31)	>33	No
Q14624	Inter-alpha-trypsin inhibitor heavy chain H4	K.IEET**T**M**T**QTPAPIQAPSAILPLPGQSVER.L	720^b, 722–723^b	$Hex_2HexNAc_2$	49	>18	Yes
Q14624	Inter-alpha-trypsin inhibitor heavy chain H4	K.IEET**T**M**T**QTPAPIQAPSAILPLPGQSVER.L	719–725^b	$Hex_3HexNAc_3$	30	>17	Yes
O95998	Interleukin-18-binding protein	D.PCP**S**QPPVFPAAK.Q	53	$Hex_2HexNAc_2$	17	>17	No
O95998	Interleukin-18-binding protein	D.PCP**S**QPPVFPAAK.Q	53	HexHexNAc	60	>33	Yes
P09603	Macrophage colony-stimulating factor 1	K.GQQPADV**T**G**T**ALPR.V	363^b, 365^b	$Hex_2HexNAc_2$	57	>20	No

TABLE I—*continued*

Uniprot/KB accession	Glycoprotein	Peptide sequence	Attachment site	Glycan	Mascot score	Mascot threshold*	ECD
P09603	Macrophage colony-stimulating factor 1	R.ISSLRPQGLSNPSTLSAQPQLSR.S	406–426[b]	HexHexNAc	55	>22	No
Q13361	Microfibrillar-associated protein 5	D.PATDETVLA.V	54[b]	HexNAc	(37)	>40	Yes
Q13361	Microfibrillar-associated protein 5	D.PATDETVLA.V	54[b]	HexHexNAc	42	>42	Yes
Q13361	Microfibrillar-associated protein 5	D.PATDETVLA.V	54[b]	HexHexNAc + Sulf	42	>38	No
Q6UXB8	Peptidase inhibitor 16	E.LQATLDHTGHTSSK.S	386–395[b]	HexHexNAc	34	>33	No
Q96FE7	Phosphoinositide-3-kinase-interacting protein 1	R.EDQTSPAPGLR.C	39[b]	HexHexNAc	51	>20	Yes
P05155	Plasma protease C1 inhibitor	K.VATTVISK.M	47–48[a]	HexHexNAc	21	>8	Yes
P05155	Plasma protease C1 inhibitor	K.VATTVISK.M	47[b], 48[a]	Hex$_2$HexNAc$_2$	16	>8	Yes
P05154	Plasma serine protease inhibitor	R.VEDLHVGATVAPSSR.R	39[b]	HexHexNAc	66	>20	Yes
P01133	Pro-epidermal growth factor	K.NQVTPLDILSK.T	801–807[b]	HexHexNAc	47	>20	No
P01133	Pro-epidermal growth factor	R.LSEPGLICPDSTPPPHLR.E	954–955[b]	HexHexNAc	57	>20	Yes
P01133	Pro-epidermal growth factor	R.LSEPGLICPDSTPPPHLR.E	954–955[b]	Hex$_2$HexNAc$_2$	72	>20	Yes
Q99075	Proheparin-binding EGF-like growth factor	D.PPTVSTDQLLPLGGGR.D	44[b]	HexHexNAc	86	>31	Yes
Q99075	Proheparin-binding EGF-like growth factor	D.PPTVSTDQLLPLGGGR.D	44[b], 47[b]	Hex$_2$HexNAc$_2$	61	>31	Yes
Q9UHG2	ProSAAS	R.GLSAASPPLAETGAPR.R	53	HexHexNAc	60	>21	Yes
Q9UHG2	ProSAAS	R.AADHDVGSELPPEGVLGALLR.V	228[b]	HexHexNAc	84	>20	Yes
Q9UHG2	ProSAAS	K.RLETPAPQVPAR.R	247	HexHexNAc	42	>17	Yes
P80370	Protein delta homolog 1	R.ALSPQQVTR.L	256[b]	HexHexNAc	30	>27	No
Q9ULI3	Protein HEG homolog 1	R.EPPPTPPRER.R	67[b]	HexHexNAc	25	>23	Yes
Q9GZM5	Protein YIPF3	K.AVAVTLQSH.-	346[b]	HexNAc	43	>19	Yes
Q9GZM5	Protein YIPF3	K.AVAVTLQSH.-	346[b]	HexHexNAc	32	>19	Yes
Q9GZM5	Protein YIPF3	K.AVAVTLQSH.-	346[b]	HexHexNAc$_2$	43	>21	Yes
Q9GZM5	Protein YIPF3	K.AVAVTLQSH.-	346[b]	Hex$_2$HexNAc$_2$	33	>20	Yes
Q9GZM5	Protein YIPF3	K.AVAVTLQSH.-	346[b]	dHexHex$_2$HexNAc$_2$	(8)	>18	Yes
Q16849	Receptor-type tyrosine-protein phosphatase-like N	K.AARPPVTPVLLE.K	441[b]	HexHexNAc	21	>20	Yes
Q4LDE5	Sushi, von Willebrand factor type A, EGF and pentraxin domain-containing protein 1	Y.DDFLDTVQETATSIGNAK.S	887–894[b]	HexHexNAc	119	>33	No
P34741	Syndecan-2	K.IPAQTKSPEETDK.E	101[b]	HexHexNAc	30[d]	>25[d]	Yes
Q6UWD8	Transmembrane protein C16orf54	M.PLTPEPPSGR.V	4[b]	HexHexNAc	30	>22	Yes
P25445	Tumor necrosis factor receptor superfamily member 6	A.QVTDINSK.G	28[b]	HexHexNAc	30	>29	Yes
Q9UFP1	Protein FAM198A	D.PGPMEPQGVTGAPATHIR.Q	53–58[b]	HexHexNAc	77	>41	No
P04070	Vitamin K-dependent protein C	G.TPAPLDSVFSSSER.A	19[b]	HexHexNAc	67	>34	Yes
P04070	Vitamin K-dependent protein C	G.TPAPLDSVFSSSER.A + Phosphorylation	19[b]	HexHexNAc	49	>39	Yes

TABLE I—*continued*

Uniprot/KB accession	Glycoprotein	Peptide sequence	Attachment site	Glycan	Mascot score	Mascot threshold*	ECD
N-linked Glycopeptides							
P02763	Alpha-1-acid glycoprotein 1	N.LVPVPITNATLDQITGK.W	33	$Hex_6HexNAc_5$	33	>17	No
P19652	Alpha-1-acid glycoprotein 2	N.LVPVPITNATLDR.I	33	$Hex_6HexNAc_5$	22	>17	No
P01009	Alpha-1-antitrypsin	R.QLAHQSNSTNIFFSPVSIATAFAMLSLGTK.A	70	$Hex_5HexNAc_4$	19	>14	No
P01009	Alpha-1-antitrypsin	K.YLGNATAIFFLPDEGK.L	271	$Hex_5HexNAc_4$	31	>15	No
P02765	Alpha-2-HS-glycoprotein	K.VCQDCPLLAPLNDTR.V	156	$Hex_5HexNAc_4$	26	>15	No
P02765	Alpha-2-HS-glycoprotein	F.NAQNNGSNFQLEEISR.A	176	$Hex_5HexNAc_4$	35	>27	No
P05090	Apolipoprotein D	R.CIQANYSLMENGK.I	65[a]	$Hex_6HexNAc_5$	36	>8	No
P05090	Apolipoprotein D	R.ADGTVNQIEGEATPVNLTEPAK.L	98	$dHex_1Hex_7HexNAc_6$	18	>10	No
Q96IY4	Carboxypeptidase B2	C.SVLLADVEDLIQQQISNDTVSPR.A	108[a]	$Hex_5HexNAc_4$	20	>19	No
P01876	Ig alpha-1 chain C region	R.PALEDLLLGSEANLTCTLTGLR.D	144	$Hex_5HexNAc_4$	31	23	No
O95998	Interleukin-18-binding protein	R.FPNFSILYWLGNGSFIEHLPGR.L	103[b]	$Hex_5HexNAc_4$	10	>4	No
O95998	Interleukin-18-binding protein	K.ALVLEQLTPALHSTNFSCVLVDPEQVVQR.H	147[b]	$dHex_1Hex_5HexNAc_4$	19	>11	No
Q96FE7	Phosphoinositide-3-kinase-interacting protein 1	R.CLNWLDAQSGLASAPVSGAGNHSYCR.N	66[b]	$dHex_1Hex_5HexNAc_4$	36	>10	No
P05155	Plasma protease C1 inhibitor	S.NPNATSSSSQDPESLQDR.G	25[a]	$Hex_5HexNAc_4$	69	>22	No
P15151	Poliovirus receptor	R.VEDEGNYTCLFVTFPQGSR.S	120[a]	$Hex_5HexNAc_4$	28	>7	No
P41222	Prostaglandin-H2 D-isomerase	K.SVVAPATDGGLNLTSTFLR.K	78	$dHex_1Hex_5HexNAc_4$	13	>6	No
P02760	Protein AMBP	K.WNITMESYVVHTNYDEYAIFLTK.K	115	$Hex_5HexNAc_4$	15	>13	No
P02760	Protein AMBP	R.YFYNGTSMACETF.Q	250[a]	$Hex_5HexNAc_4$	21	>13	No
P00734	Prothrombin	R.GHVNITR.S	121[b]	$Hex_5HexNAc_4$	19	>4	Yes
P00734	Prothrombin	R.YPHKPEINSTTHPGADLQENFCR.N	143[b]	$Hex_5HexNAc_4$	30[d]	>18[d]	Yes
P07911	Uromodulin	R.CNTAAPMWLNGTHPSSDEGIVSR.K	232[a]	$Hex_7HexNAc_6$	17	>15	No
P07911	Uromodulin	K.QDFNITDISLLEHR.L	322[a]	$dHex_1Hex_5HexNAc_6$	13	>10	No
P07911	Uromodulin	R.NETHATYSNTLY.L	396[a]	$Hex_6HexNAc_5$	19	>19	No
Q6EMK4	Vasorin	R.LHEITNETFR.G	117[a]	$Hex_5HexNAc_4$	(11)[d]	>15[d]	Yes
P25311	Zinc-alpha-2-glycoprotein	R.FGCEIENNR.S	128[a]	$Hex_5HexNAc_4$	14	>11	No

[a] Site occupancy reported in the UniProtKB/Swiss-Prot database, glycan unknown.
[b] Site occupancy and glycan not reported in the UniProtKB/Swiss-Prot database.
[c] For N-glycan microheterogeneity, see supplementary Table III.
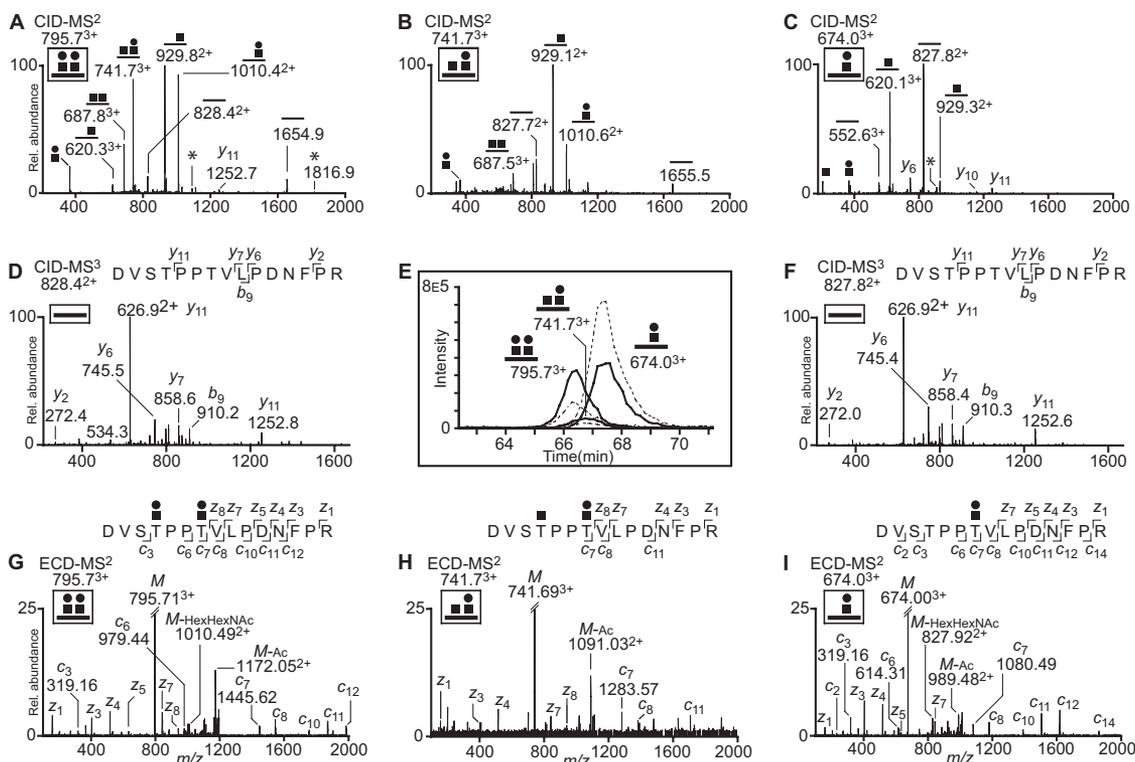[d] Mascot score and threshold values obtained for ECD data.

FIG. 2. **LTQ-FTICR mass spectrometry of urinary *O*-linked glycopeptides derived from Insulin-like growth factor II.** *A*, CID-MS^2 spectrum of the Hex$_2$HexNAc$_2$ glycoform (*m/z* 795.7093^{3+}). *B*, CID-MS^2 spectrum of the HexHexNAc$_2$ glycoform (*m/z* 741.6934^{3+}). *C*, CID-MS^2 spectrum of the HexHexNAc glycoform (*m/z* 673.9967^{3+}). *D*, CID-MS^3 spectrum of the unmodified peptide at *m/z* 828.4^{2+} from panel *A*. *E*, Extracted base peak chromatograms showing the elution profile and intensity of triply charged (solid line) and doubly charged (dashed line) parent ions. *F*, CID-MS^3 spectrum of the unmodified peptide at *m/z* 827.8^{2+} from panel *C*. *G*, ECD-MS^2 spectrum of the triply charged Hex$_2$HexNAc$_2$ glycoform (*m/z* 795.7085^{3+}) showing the effective dissociation of the precursor, which revealed the novel glycosylation site at Thr[96]. *H*, ECD-MS^2 spectrum of the triply charged HexHexNAc$_2$ glycoform (*m/z* 741.6914^{3+}). *I*, ECD-MS^2 spectrum of the triply charged HexHexNAc glycoform (*m/z* 673.9979^{3+}), which confirms the previously reported glycosylation site at Thr[99]. The isolated ions subjected to CID-MS^n/ECD-MS^2 fragmentation are boxed and schematically illustrated in each panel. Circle, Hex; square, HexNAc; bold line, D[93]VS*TPP*T*VLPDNFPR[107] peptide. Potential hexose rearrangements products are depicted with asterisk.

(see below). The Y-type fragment ion at *m/z* 687.5 corresponding to [peptide+HexNAc$_2$+3H]$^{3+}$ showed that two HexNAc residues were attached to the peptide but did not reveal if they were located on individual Ser/Thr or linked in a core 2-like manner. Again, a diagnostic [HexNAcHexNAc+H]$^+$ ion at *m/z* 407 was not observed, indicating that the HexNAc residues were located on separate Ser/Thr residues. Approximately 1 min later the HexHexNAc glycoform eluted (*m/z* 674.0 in Fig. 2*E*) and the CID-MS^2 spectrum (Fig. 2*C*) showed intense Y-ions corresponding to the loss of Hex (*m/z* 620.1 and *m/z* 929.3) and HexHexNAc (*m/z* 552.6 and *m/z* 827.8) from precursor. CID-MS^3 fragmentation of the peptide ion (Y$_0$-ion) at *m/z* 828.4 (Fig. 2*A*) and *m/z* 827.8 (Fig. 2*C*) resulted in *b*- and *y*-ions, shown in Figs. 2*D* and 2*F*, which were used for peptide identification through the Mascot algorithm. The CID-MS^3 spectra of the Y$_0$-ions in Figs. 2*A*–2*C* (*m/z* 828) were all matched to the tryptic D[93]VSTPPTVLPDNFPR[107] peptide of Insulin-like growth factor II (IGF-II, UniProt/KB accession P01344) with ion scores of 67 (*p* < 0.05 threshold; >26), 39 (*p* < 0.05 threshold; >21) and 63 (*p* < 0.05 threshold; >21) for

the Hex$_2$HexNAc$_2$ (Fig. 2*D*), HexHexNAc$_2$ (not shown), and HexHexNAc (Fig. 2*F*) glycoform, respectively (Table I).

Assignment of Glycan Attachment Sites by ECD—We also acquired ECD-MS^2 spectra of the triply charged D[93]VSTPPTVLPDNFPR[107] glycopeptides from IGF-II, with Hex$_2$HexNAc$_2$ (Fig. 2*G*), HexHexNAc$_2$ (Fig. 2*H*) and HexHexNAc (Fig. 2*I*) glycans. Fragmentation of triply charged precursors generated sufficient *c*- and *z*-ions to be used for glycosylation site identification purposes. For the ECD-MS^2 of the triply charged Hex$_2$HexNAc$_2$ glycoform (Fig. 2*G*), the c_3-ion was observed without the additional mass of any glycan (*m/z* 319.16) indicating that Ser[95] was not modified. The c_7-ion, however, was detected with the additional mass of Hex$_2$HexNAc$_2$ (*m/z* 1445.62) showing that the glycan(s) had to reside within the Thr[96]-Pro-Pro-Thr[99] sequence. The cyclic structure of proline precludes ECD induced N-terminal cleavage and c_4, c_5, z_{10}, and z_{11} ions can thus not be observed. The only fragment ions that can resolve the glycan attachment site(s) are therefore z_9 and c_6. A glycosylated c_6 fragment was indeed observed at *m/z* 979.44 (Fig. 2*G*), which showed that
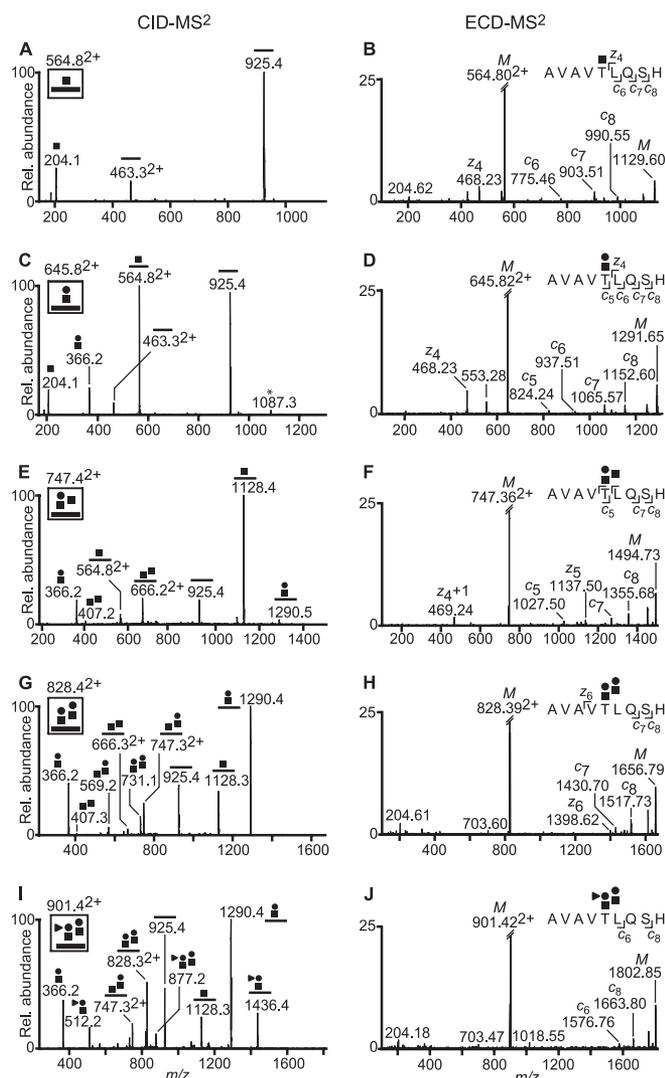
FIG. 3. **Microheterogeneity of the A³⁴²VAV*T*LQSH³⁵⁰ *O*-linked glycopeptide from urinary protein YIPF3.** (*A*) CID-MS² (m/z 564.7991²⁺) and (*B*) ECD-MS² spectra (m/z 564.7987²⁺) of the HexNAc glycoform which pinpoint the novel glycosylation site to Thr³⁴⁶. (*C*) CID-MS² spectrum (m/z 645.8257²⁺) and (*D*) ECD-MS² spectrum (m/z 645.8252²⁺) of the HexHexNAc glycoform. (*E*) CID-MS² spectrum of the HexHexNAc₂ glycoform (m/z 747.3658²⁺) with a diagnostic ion at *m/z* 407, indicating a branched core 2-like structure and (*F*) ECD-MS² spectrum (m/z 747.3655²⁺) for the same HexHexNAc₂ glycoform showing that the entire glycan moiety resides on Thr³⁴⁶. (*G*) CID-MS² spectrum of the Hex₂HexNAc₂ glycoform (m/z 828.3925²⁺) with oxonium fragments ions at *m/z* 407 and *m/z* 569. (*H*) ECD-MS² spectrum of the Hex₂HexNAc₂ glycoform (m/z 828.3924²⁺). (*I*) CID-MS² spectrum of the dHexHex₂HexNAc₂ glycoform (m/z 901.4224²⁺) which shows a complex glycosidic fragmentation pattern and (*J*) ECD-MS² of the same fucosylated glycoform (m/z 901.4221²⁺) showing once again that the entire glycan moiety is attached to Thr³⁴⁶. The isolated ions subjected to CID-MSⁿ/ECD-MS² fragmentation are boxed and schematically illustrated in each panel. Triangle: dHex; circle, Hex; square, HexNAc; bold line, A³⁴²VAV*T*LQSH³⁵⁰ peptide. Potential hexose rearrangements products are depicted with asterisk.

Thr⁹⁶ harbored a single HexHexNAc. The c_7 fragment was observed at *m/z* 1445.62, which mapped the second HexHexNAc to Thr⁹⁹. The glycan sequence, determined as two separate HexHexNAc-*O*-Ser/Thr structures by CID-MS² (Fig. 2*A*), was thus mapped by ECD-MS² (Fig. 2*G*) to two individual amino acids, *i.e.* Thr⁹⁶ and Thr⁹⁹ of IGF-II. The ECD-MS² spectrum of the triply charged HexHexNAc₂ glycoform (Fig. 2*H*) allowed us to verify the peptide sequence and the presence of a HexHexNAc₂ moiety within the Asp⁹³-Val⁹⁴-Ser⁹⁵-Thr⁹⁶-Pro⁹⁷-Pro⁹⁸-Thr⁹⁹ region. However, we did not detect any fragment ions that could differentiate whether Thr⁹⁶ or Thr⁹⁹ was modified with the single HexNAc. For the HexHexNAc glycoform (Fig. 2*I*) the c_3 ion was once again observed without the additional mass of the carbohydrate, showing that Ser⁹⁵ was not modified. Furthermore, the c_6 was detected at *m/z* 614.31 and was thus not glycosylated and showed that Thr⁹⁶ was not the glycosylation site. In contrast, the c_7-ion was detected with the additional mass of HexHexNAc (365.13 Da) at *m/z* 1080.49, thereby pinpointing the glycosylation site to Thr⁹⁹ of IGF-II as previously described (52). Taken together, these experiments also revealed the site occupancy (macroheterogeneity) within the D⁹³VS*T*PP*T*VLPDNFPR¹⁰⁷ tryptic glycopeptide, *i.e.* the initial HexHexNAc glycosylation occurs at Thr⁹⁹ whereas the second HexHexNAc is attached to Thr⁹⁶.

Fragments corresponding to the loss of 43.02 Da from glycopeptide precursors were also observed in ECD-MS², seen at *m/z* 1172.05 (Fig. 2G) and at *m/z* 989.48 (Fig. 2*I*). A plausible explanation for these secondary fragments has been attributed to the loss of an acetyl radical ($C_2H_3O^\bullet$) from the *N*-acetyl moiety of HexNAc containing glycopeptides (53). Also, elimination of HexHexNAc from precursor ions was occasionally observed in ECD-MS² (*m/z* 1010.49 in Fig. 2*G* and *m/z* 827.92 in Fig. 2*I*) but such fragmentation channels were minor dissociation pathways, which did not have a negative impact on the interpretation of ECD spectra. In total, 32 *O*-linked glycosylation sites were manually assigned to unique Ser/Thr residues using ECD (Table I and supplemental Fig. S5). We defined 8 *O*-glycan attachment sites by CID and in total we thus identified 40 unique *O*-glycosylation sites.

O-linked Glycopeptide Microheterogeneity and Modifications—In a few instances, also other glycoforms apart from the HexNAc-*O*-Ser/Thr structure were identified (Table I, supplemental Table S2 and supplemental Fig. S5). For the C-terminal tryptic peptide A³⁴²VAV*T*LQSH³⁵⁰ from protein YIPF3 a single HexNAc, in accordance with the Tn-antigen, (GalNAcα-*O*-Ser/Thr, Fig. 3*A*) was identified. The ECD-MS² spectrum showed that the HexNAc was attached to the Thr³⁴⁶ residue (Fig. 3*B*). The HexHexNAc glycoform was also identified by CID-MS² (Fig. 3*C*) and ECD-MS² (Fig. 3*D*). Further, three core 2-like structures with Hex(HexNAc)HexNAc (Fig. 3*E*), Hex(HexHexNAc)HexNAc (Fig. 3*G*) and dHexHex-(HexHexNAc)HexNAc (Fig. 3*I*) glycans were also identified. One glycosylation site on the Thr³⁴⁶ residue was mapped for these *O*-linked glycopeptides by ECD-MS² (Fig. 3*F*, 3*H*, and

3*J*). The presence of the HexNAcHexNAc B/Y-type ion (*m/z* 407, Fig. 3*E* and 3*G*); the HexHexNAc$_2$ B/Y-type ion (*m/z* 569, Fig. 3*G*); Hex$_2$HexNAc$_2$ (*m/z* 731, Fig. 3*G*) and dHexHex$_2$HexNAc$_2$ (*m/z* 877, Fig 3*I*) verified that these glycans exceeded the HexHexNAc structure in complexity and thus confirmed the presence of one as opposed to two glycosylation sites for this peptide. B/Y-type oxonium ions exceeding *m/z* 407, *e.g.* at *m/z* 569 equally well matched ions corresponding to [Hex-(HexNAc)-HexNAc + H]$^+$ and [HexNAc-Hex-HexNAc +H]$^+$, *i.e.* a branched or a linear glycan sequence, respectively. Thus, B/Y-type ions at *m/z* 569 were unable to differentiate core 2-like glycans from elongated (linear) core 1-like structures. The same limitation is true for B-type ions at m/z 731 (Fig. 3*G*), corresponding to the entire Hex$_2$HexNAc$_2$ moiety of *O*-linked glycopeptides. Y-type oxonium ions at *m/z* 528, corresponding to [Hex-HexNAc-Hex + H]$^+$, could potentially reveal a linear *O*-glycan sequence but such ions were not observed in any CID-MS^n experiments for Hex$_2$HexNAc$_2$ glycoforms in this study.

Additionally, we identified secondary modifications of some *O*-linked glycopeptides. The CID-MS^2 fragmentation spectrum of the HexHexNAc glycosylated P^{52}A*T*DETVLA^{60} peptide (Microfibrillar-associated protein 5, UniProt/KB accession Q13361) (Fig. 4*A*) showed an initial loss of ~80 Da (*m/z* 641.0), which we tentatively assigned as a sulfate group (79.9568 Da), but which could in theory also be a phosphate group (79.9663 Da). The precursor ion (*m/z* 681.2804^{2+}, not shown) was found to deviate by 1.69 ppm (–5.28 ppm for a phosphorylated precursor ion) from the theoretical monoisotopic mass of a sulfated precursor ion. In addition to the oxonium ions at *m/z* 204 (HexNAc) and *m/z* 366 (HexHexNAc), a fragment ion at *m/z* 446 was also observed which indicated that the sulfate group resides on the glycan and not on the peptide (Fig. 4*A* and Fig. 4*B*). Co-eluting with the sulfated precursor, we also observed the nonsulfated glycoform, *i.e.* the HexHexNAc modified P^{52}A*T*DELVLA^{60} peptide, which was also characterized by CID-MS^n and ECD-MS^2 fragmentation (supplemental Fig. S5). The FTICR-MS^1 measured mass difference between the sulfated (*m/z* 681.3019^{2+}) and nonsulfated (*m/z* 641.3019^{2+}) variants of the HexHexNAc glycosylated P^{52}A*T*DETVLA^{60} peptide was found to be 79.9570 Da, which deviates from the theoretical value of a sulfate group (79.9568 Da) only by 0.0002 Da. Although the *m/z* 446 ion, corresponding to HexHexNAc+Sulf, was detected and mass measured in the ion trap, the accurate mass of the sulfate group was thus indirectly confirmed by the mass measurements of the precursor ions in the ICR cell. Unfortunately, whether the Hex or HexNAc was carrying the secondary modification could not be defined.

The CID-MS^2 spectrum of the HexHexNAc glycosylated *T*^{19}PAPLDSVFSSSER^{32} peptide (Vitamin K-dependent protein C, UniProt/KB accession P04070) is shown in Fig. 4*C*. This glycopeptide was also detected with a mass increment of ~80 Da. However, the CID-MS^2 fragmentation of this glyco-
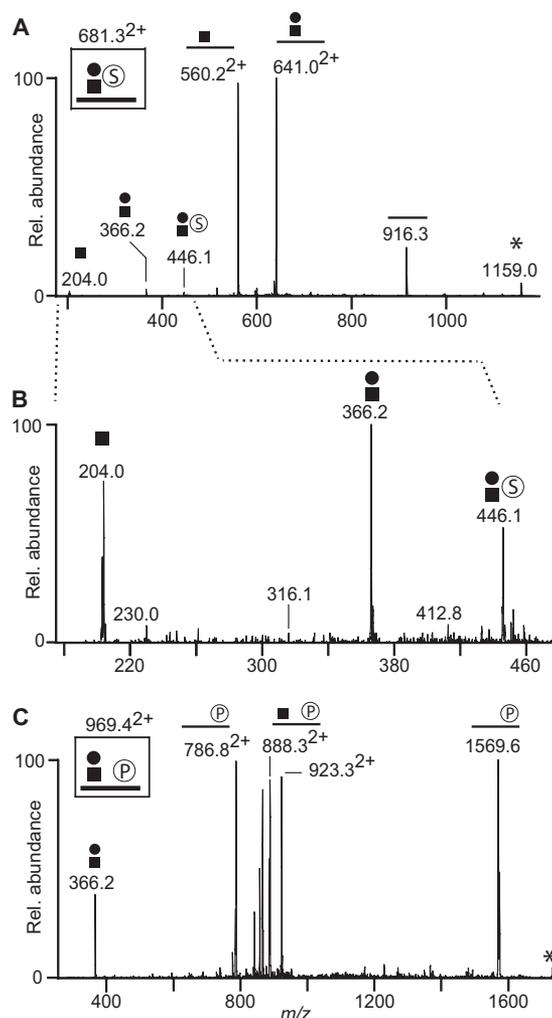

Fig. 4. **Modifications of *O*-linked glycopeptides.** *A*, CID-MS^2 spectrum (*m/z* 681.2804^{2+}) of P^{52}A*T*DETVLA^{60} peptide (Microfibrillar-associated protein 5) with a tentative sulfate group on the HexHexNAc component. *B*, Expansion in the low mass range (*m/z* 180–460) showing the oxonium fragment ions from panel *A*. *C*, CID-MS^2 spectrum (*m/z* 969.4200^{2+}) of the HexHexNAc glycosylated *T*^{19}PAPLDSVFSSSER^{32} peptide with a tentative phosphate group attached to the peptide. The isolated ions subjected to CID-MS^n fragmentation are boxed and schematically illustrated in each panel. Circle, Hex; square, HexNAc; circled S, sulfate; circled P, phosphate; bold line, peptide. Potential hexose rearrangements products are depicted with asterisk.

peptide resulted in an initial loss of Hex (to *m/z* 888.3) followed by a loss of HexNAc (to *m/z* 786.8), showing that the modification, tentatively assigned as a phosphorylation, was attached to the peptide and not to the glycan. The precursor ion (*m/z* 969.4200^{2+}, supplemental Fig. S5) was found to deviate by 3.30 ppm (8.20 ppm for a sulfated precursor ion) from the theoretical monoisotopic mass of a phosphorylated precursor ion. The results in Fig. 4*C* indicate that *O*-linked glycans are more susceptible to CID-induced fragmentation by comparison to phosphate groups. ECD-MS^2 fragmentation (supplemental Fig. S5) allowed us to pinpoint the HexHexNAc-*O*-

FIG. 5. **LTQ-FTICR mass spectrometry of urinary *N*-linked glycopeptides.** *A*, MS[1] acquisition of individual peptide glycoforms for the ADGTVNQIEGEATPVN[98]LTEPAK peptide from apolipoprotein D. Mass accuracy and resolution is shown in the insert. *B*, CID-MS[2] spectrum of the bi-antennary *N*-linked glycopeptide from apolipoprotein D (m/z 1292.9104[3+]). *C*, CID-MS[3] spectrum of the peptide+HexNAc fragment used for peptide identification by the Mascot algorithm. Peptide *b*-ions are omitted for clarity. *D*, CID-MS[2] spectrum of the fucosylated tetra-antennary *N*-linked glycopeptide (QDFN[322]ITDISLLEHR) from uromodulin (m/z 1394.9225[3+]). *E*, Partially annotated CID-MS[3] spectrum at m/z 1909.8 from panel *D*. *F*, CID-MS[3] spectrum of the peptide+dHexHexNAc fragment from uromodulin used for Mascot identification. *G*, CID-MS[2] spectrum of the pentuply charged bi-antennary *N*-linked glycopeptide (YPHKPEIN[143]STTHPGADLQENFCR) from prothrombin (m/z 867.5770[5+]). *H*, CID-MS[3] spectrum at m/z 993.3 from panel *G*. *I*, ECD-MS[2] spectrum of the same *N*-linked glycopeptide from prothrombin (867.5761[5+]). The isolated ions subjected to CID-MS[n]/ECD-MS[2] fragmentation are boxed and schematically illustrated in each panel. Triangle, dHex; circle, Hex; square, HexNAc; bold line, peptide. Pyroglutamate is noted as pyQ in panel *F*.

sequence to Thr[19] but the phosphorylated serine residue, among the four possible, was not identified (Table I and supplemental Fig. S5).

CID- and ECD-fragmentation of N-linked Glycopeptides—Fifty-eight glycopeptides, corresponding to 25 differently *N*-glycosylated peptides from 17 urinary glycoproteins were identified (0.0% false positive identifications) in the formic acid released glycopeptide fractions (supplemental Fig. S1C and supplemental Fig. S6). They are all listed together with their *N*-linked glycans, their attachment sites and Mascot scores of the dominating glycopeptides in Table I and in supplemental Table S3. As a general feature, we observed the presence of several glycoforms for each *N*-linked glycopeptide. The relative abundance of specific glycoforms was determined by integrating chromatographic peaks for individual peptide glycoforms and the values were used to estimate the relative distribution of *N*-glycan microheterogeneity at each site. Oligosaccharide composition corresponding to the bian-

tennary complex type structure was typically dominating, although triantennary and fucosylated bi- and triantennary glycoforms were also identified (supplemental Table S3). Sialic acid micro-heterogeneity was not observable since sialic acids were hydrolyzed in the preparative procedure. To illustrate the used methodology, MS[n] of *N*-linked glycopeptides originating from three well-known *N*-glycoproteins (apolipoprotein D, uromodulin and prothrombin) are described in more detail (Fig. 5). Firstly, the FTICR-MS[1] spectrum (Fig. 5*A*) showed the ADGTVNQIEGEATPVN[98]LTEPAK peptide from Apolipoprotein D (UniProt/KB accession P05090) with *N*-linked glycans corresponding to the complex type biantennary and fucosylated bi-, tri-, and tetraantennary structures. The *N*-linked glycan structures were all deduced by CID-MS[n] fragmentation of selected precursor ions. The CID-MS[2] spectrum of the biantennary *N*-linked glycopeptide at m/z 1292.9 (Fig. 5*B*) resulted in intense charge reduced fragment ions at m/z 1756.7 and m/z 1675.5 corresponding to the loss of HexHex-

NAc and Hex$_2$HexNAc, respectively. The third most intense ion (*m/z* 1229.5) resulted from a glycosidic cleavage at the GlcNAcGlcNAc chitobiose core and corresponds to the [peptide+HexNAc+2H]$^{2+}$ (Y$_1$) ion. CID-MS^3 of the [peptide+HexNAc+2H]$^{2+}$ ion (Fig. 5*C*) induced peptide backbone fragmentation into *b*- and *y*-ions and were used for identification of the glycan attachment site and peptide sequence by the Mascot algorithm.

Second, the CID-MS^2 fragmentation of a precursor at *m/z* 1394.9, corresponded to a fucosylated tetra-antennary complex type *N*-glycopeptide from uromodulin (UniProt/KB accession P07911) and resulted in a prominent charge reduced fragment ion at *m/z* 1909.8 because of the loss of a terminal HexHexNAc moiety and a proton (Fig. 5*D*). The second most intense fragment (*m/z* 1017.4) corresponded to [peptide+dHexHexNAc+2H]$^{2+}$, indicating that the fucose resided on the asparagine linked GlcNAc. Additional fragment ions were visible at *m/z* 1836.8, *m/z* 1727.6, and *m/z* 1646.7 corresponding to the loss of dHexHexHexNAc, Hex$_2$HexNAc$_2$ and Hex$_3$HexNAc$_2$, respectively, and revealed partial structural information on the *N*-linked glycan. The CID-MS^3 spectrum at *m/z* 1909.8 (Fig. 5*E*) showed further sequential glycosidic fragmentation and the entire *N*-glycan sequence was verified. Ideally, the fragment ion corresponding to [peptide+HexNAc+2H]$^{2+}$ at *m/z* 944.6 (Fig. 5*D*) would have been used for the peptide identification but because of its low abundance it was not selected for CID-MS^3 fragmentation. Low abundance of Y$_1$-ion peaks was found to be a common feature for core fucosylated *N*-glycopeptides in CID-MS^2 spectra (supplemental Fig. S6). Instead, the fragment ion corresponding to [peptide+dHexHexNAc+2H]$^{2+}$ (*m/z* 1017.4, Fig. 5*D*) was selected for CID-MS^3 fragmentation (Fig. 5*F*). We observed an intense peak at *m/z* 943.9 corresponding to the loss of dHex together with minor peaks corresponding to peptide fragmentation and the MS^3 spectrum was matched to the tryptic QDFN^{322}ITDISLLEHR peptide of uromodulin, with a Mascot score of 13 (*p* < 0.05 threshold; >10).

Third, the CID-MS^2 fragmentation of a pentuply charged biantennary *N*-linked glycopeptide at *m/z* 867.6 (Fig. 5*G*) rendered in a different fragmentation pattern compared with a triply charged biantennary *N*-glycopeptide (compare Figs. 5*B* and 5*G*) because of the different charge states, 3+ *versus* 5+. For the pentuply charged precursor we observed abundant glycosidic fragmentation of the terminal HexHexNAc residues and no apparent ion intensity corresponding to the peptide+HexNAc fragment. Subsequent CID-MS^3 at *m/z* 993.3 in (Fig. 5H) allowed for verification of the biantennary glycan structure but the amino acid sequence remained unidentified because of the lack of CID-MS^3 data on the peptide+HexNAc fragment ion. However, considering the high charge state, and thus the relatively low *m/z* ratio, this glycopeptide was efficiently fragmented into *c*- and *z*-type ions by ECD-MS^2 (Fig. 5*I*) and the peptide sequence was identified to originate from the tryptic YPHK-

PEIN^{143}STTHPGADLQENFCR peptide from prothrombin (UniProtKB accession P00734). The combination of CID-MS^n with ECD-MS^2 was found to be useful in the identification of an additional *N*-linked glycopeptide (supplemental Fig. S6), namely the tryptic LHEITN^{117}ETFR peptide of vasorin (UniProt/KB accession Q6EMK4).

DISCUSSION

The production of urine takes place in the nephron and involves a complex process of ultrafiltration, reabsorption and secretion, eventually leading to the formation of a complex solution containing metabolic waste products, proteins and peptides (54). The high content of salt and metabolic waste products in human urine requires sample purification for the removal of interfering compounds and isolation of urinary proteins prior to proteomic analysis. As yet, there is no universal method that offers complete recovery of the urinary proteome. Various approaches have been investigated for this purpose with each method offering advantages and disadvantages when compared with each other (55). In our study, the choice of sample preparation method was important not only for qualitative recovery of urinary proteins but was also essential for our downstream application, *i.e.* mild periodic acid oxidation of sialic acids. Efficient and selective oxidation of sialic acids was critical for the enrichment procedure of urinary glycoproteins, a reaction conducted under mild conditions employing only 2 mM periodic acid. Thus, the sample preparation method had to offer qualitative recovery of the urinary proteome and deplete metabolic waste products that might interfere or quench the subsequent oxidation of sialic acids. Several sample preparation methods were examined for this purpose, including organic solvent precipitation (acetone and trichloroacetic acid), spin column purification, size exclusion and reversed phase (C18) chromatography (not shown). Unfortunately, all were found to yield inadequate sample purity and failed in removing residual urinary pigments, which interfered with the sialic acid oxidation.

Eventually, we explored dialysis followed by lyophilization as a way to isolate and concentrate urinary proteins in a two-step procedure. Dialysis of urine against water alone was inefficient (supplemental Fig. S2) but the addition of 1.5% SDS and dialysis at 60 °C was found to yield sufficient sample purity for subsequent sialic acid oxidation. The dilute dialysates were subsequently concentrated through lyophilization to minimize the risk of unnecessary sample losses. Albeit time consuming, the preparative procedure employed in this study was thus justified by the strict requirement of sample purity and qualitative protein recovery.

Given that the dialyzed samples would serve as the basis for enrichment of sialoglycoproteins, it was also important to validate the preparative procedure to ensure that a representative urinary proteome was isolated following dialysis and lyophilization. By comparing our data set with the comprehensive proteomic studies of Adachi *et al.* and Kentsis *et al.*

(8, 9), we concluded that 90% of our protein identifications showed a nearly uniform overlap with the data sets of these studies (supplemental Fig. 3*A*). This observation confirmed that the glycoproteomic data would not mirror an atypical urinary subproteome as a result of the preparative procedure. It should be stressed that our proteomic analysis was not intended to expand the urinary proteome coverage. Thus, in contrast to previous studies, we did not deplete or prefractionate the urine sample prior to the one-dimensional electrophoretic separation, which may explain the relatively low number of protein identifications in this study.

Subsequent enrichment of sialoglycoproteins from the dialysates was achieved through conjugation of oxidized sialic acids to hydrazide beads (supplemental Fig. S1B and S1C). Although side reactions with terminal Hex or HexNAc residues of nonsialylated glycoproteins cannot be completely avoided, the mild oxidation constitutes the first step of introducing specificity to the enrichment procedure. Under these mild conditions, oxidation takes place primarily at the glycerol side chain (C7-C9) of sialic acids. In other words, hydrazide reactive aldehyde groups are specifically introduced on sialic acid by periodic acid oxidation at 0 °C. Consequently, targeted enrichment of sialoglycoproteins is enabled by reducing sample complexity through sequential washes of the solid phase to remove nonglycosylated and nonsialylated urinary proteins.

Following trypsin digestion and peptide extraction, the solid phase was extensively washed to remove any remaining nonglycosylated peptides in order to avoid interference by *e.g.* ion suppression effects in downstream analyses. The covalently linked glycopeptides were subsequently released by mild formic acid hydrolysis for MS-analysis. The formic acid treatment results in specific hydrolysis of sialic acid glycosidic bonds without affecting linkages between dHex, Hex or HexNAc residues, and thereby represents the second step of specificity in the glycopeptide enrichment procedure. Only species sensitive to formic acid cleavage are released from the hydrazide beads, which includes glycopeptides conjugated through sialic acids and exclude nonsialylated glycopeptides. Thus, other biomolecules harboring hydrazide reactive groups but lacking formic acid sensitive linkages are also excluded in this step. The combination of both specificity steps, *i.e.* mild periodic acid oxidation and mild formic acid hydrolysis, thus allows for selective isolation of desialylated glycopeptides. Consistent with this statement, base peak chromatograms of formic acid released fractions revealed various *N*- and *O*-linked glycopeptides as the dominating components (supplemental Fig. S4) with >80% of the subsequent CID-MS^2 spectra possessing typical glycopeptide fragmentation patterns accompanied by diagnostic carbohydrate oxonium ions (56).

Identification of glycan- and peptide sequences was enabled by subjecting enriched glycopeptides to multiple rounds of CID fragmentation. CID-MS^2 spectra of HexHex-NAc glycoforms displayed prominent Y_1 and Y_0 fragments that were used to identify HexHexNAc-*O*-Ser/Thr sequences. Weak fragment ions corresponding to the mass of peptide+Hex, indicated with an asterisk in Figs. 2 to 4, were also observed during CID-MS^2. These observations may contradict the HexHexNAc-*O*-Ser/Thr sequence outlined above, suggesting a Hex residue as the internal peptide linked monosaccharide. However, migration of hexose residues upon CID of protonated *N*-glycans and *N*-glycopeptides has been previously observed (57, 58), resulting in fragment ions which may lead to incorrect structural predictions. We speculate that the weak peptide+Hex fragment ions generated upon CID of protonated *O*-linked glycopeptides are most likely caused by hexose migrations similar to those observed for protonated *N*-linked glycopeptides but further studies are needed to verify these findings.

O-linked glycopeptides containing the $Hex_2HexNAc_2$ glycoform generally required five CID-MS^3 experiments to delineate glycan- and peptide sequences. For *O*-linked glycopeptides with more than four monosaccharide units, isolation of intact peptide ions for CID-MS^3 fragmentation proved difficult because of the increasing dominance of glycosidic fragments in MS^2 spectra. Thus, the characterization of glycan- and peptide sequences for *O*-linked glycopeptides glycosylated beyond the simple core 1-like structure was rapidly complicated by the increasing number of monosaccharides. This is in contrast to *N*-linked glycopeptides which are readily identified even though they contain 9–13 monosaccharide units. Difficulties in characterizing *O*-linked glycopeptides with the $Hex_2HexNAc_2$ glycoform arise not only from isolation of Y_0-ions for CID-MS^3, but also from assigning the correct glycan sequence for the carbohydrate moiety. Y-type fragments are usually unable to resolve complex *O*-glycan sequences since they are equally well matched to the fragmentation pattern of different glycoforms. Thus, the identification procedure for *O*-glycopeptides is not easily automated and careful manual annotation is still necessary for correct assignment of glycan sequences.

By combining the CID and ECD data for each precursor ion, complementary information of core glycosylation could be gathered. ECD induced peptide fragmentation of $Hex_2HexNAc_2$ glycoforms revealed if the oligosaccharide components were located on two separate amino acids, suggesting a macroheterogeneity with two core 1-like glycans (Fig. 2), or different glycans on one single amino acid, indicating site-specific microheterogeneity (Fig. 3). However, ECD fragmentation does not provide structural information on the glycan sequence *per se* and determination of glycan sequence was therefore mainly based on CID-MS^n data. Thus, the main purpose of the ECD experiments was to determine the amino acid attachment sites of *O*-linked glycans. Traditionally, *O*-linked glycans are attached to serine or threonine residues but recently we reported a tyrosine residue to be modified by a sialylated *O*-linked glycan on amyloid beta peptides in human cerebrospinal fluid (59). However, our ECD

experiments did not reveal any tyrosine glycosylated peptides in the urine samples, suggesting that complex tyrosine glycosylation is rare, and possibly more tissue specific, than mucintype *O*-glycosylation on the serine and threonine residues.

The majority of *O*-linked glycopeptides in Table I were thus identified with a single core 1-like glycan, which raises the issue of whether or not proteins *O*-glycosylated with core-1 like glycans are positively selected for by our approach. We argue that terminal sialic acids should be equally well oxidized by the periodic acid treatment, regardless of their core glycan structure, and that *O*-glycopeptides are equally well enriched on the hydrazide beads, given that they are sialylated to the same extent. The release mechanism should also not be dependent on the core glycan structure but only related to the hydrolysis of acid sensitive NeuAc-Gal or NeuAc-GalNAc glycosidic linkages. The subsequent detection of glycopeptides in LC-FTICR-MS[1] is largely dependent on two factors: 1) the chromatographic properties of the peptide backbone, *i.e.* only glycopeptides of suitable length and hydrophobic character will be resolved by the C18 column; and 2) the physiochemical properties of the peptide backbone, which will dictate the extent of ionization and the stability of the parent ions. *O*-glycosylation microheterogeneity was found to have a minor impact on chromatographic retention times (Fig. 2*E*) with various peptide glycoforms eluting within a narrow time frame. The chromatography is thus not expected to favor any particular peptide glycoform since the retaining properties of the C18 column are generally dependent on the peptide composition rather than on the glycan structure. Thus, enrichment and characterization of *O*-glycan microheterogeneity, *i.e.* core 1-like *versus* core 2-like glycosylations, is probably not limited by the chromatographic resolution since different core glycans attached to the same peptide backbone are expected to be resolved equally well. Positive mode ionization of glycopeptides results in detection of $[M+nH]^{n+}$ molecular ions, an outcome that is dependent on the proton affinity of the peptide backbone. This property justifies the comparison of signal intensities not only for detection of microheterogeneity but also for relative quantification of individual peptide glycoforms (60). We were also able to observe extensive microheterogeneity for specific *O*-glycopeptides, as demonstrated for the $A^{342}VAVTLQSH^{350}$ peptide of protein YIPF3 in Fig. 3. This *O*-glycopeptide was identified in five different core glycoforms ranging from a single HexNAc residue to a fucose containing pentasaccharide, clearly showing that our approach is not selective for *O*-glycopeptides occupied only by core 1-like glycans. Taken together, this indicates that the observed HexHexNAc core 1-like glycans are indeed the predominant *O*-glycans of the sialylated human urinary glycoproteome. In an earlier study the sialylated core 1 glycan was really shown to be the dominating *O*-glycan for uromodulin in nonpregnant female and male urine samples whereas Lewis structures on *O*-glycans were typical for uromodulin in pregnant female urine (61). We were unable to identify any *O*-linked glycopep-

tides from uromodulin in our study, which suggests that the *O*-glycans of uromodulin are located within trypsin-inaccessible regions of the protein. Alternatively, the trypsin digestion might also result in short, hydrophilic O-glycopeptides which were not retained by the C18 column and thus not detected during analysis. This limitation, which extends to all urinary glycoproteins and is valid for both *N*- and *O*-linked glycosylations, may be circumvented by the use of alternative proteases.

Several urinary glycoproteins, *e.g.* CD44, macrophage colony-stimulating factor 1, vasorin, complement component 7 and protein HEG homolog, identified as enriched glycopeptides in Table I, are each estimated to constitute less than 0.1–0.02% (by mass) of the core urinary proteome (10). This clearly shows that sialylated glycoproteins present in minute amounts in the urine are selectively made accessible for glycoproteomic characterization by the enrichment procedure. Notably, several other glycoproteins of Table I have been identified as potential biomarkers, *e.g.* elevated levels of urinary IGF-2 in urothelial carcinoma of the bladder (62) and it is not unlikely that these changes are accompanied by aberrant *O*-glycan profiles. The sialyl-Tn antigen (Neu5Acα2–6GalNAcα-*O*-Ser/Thr) is a rare glycoepitope in normal tissue but high expression levels are known to occur in ovarian (63), gastric (64), colorectal (65) and pancreatic (66) carcinomas. Existing evidence also indicates that *O*-glycan occupancy is increased in cancer cells (67, 68). The ability to probe both these features simultaneously, *i.e.* site occupancy and *O*-glycan microheterogeneity, thus offers a unique opportunity to link aberrant glycans with distinct proteins. Although nonsialylated structures, *e.g.* Tn-antigen (GalNAcα-*O*-Ser/Thr) or high-mannose type *N*-glycans are not enriched by the procedure, this analytical strategy could provide further insight into the process of pathogenesis for a wide range of diseases by identifying key proteins that are aberrantly glycosylated. Thus, the methodology and the results presented in this study should be of value for further exploration of the urinary glycoproteome in search of novel disease biomarkers.

§ To whom correspondence should be addressed: Department of Clinical Chemistry and Transfusion Medicine, Institute of Biomedicine, Sahlgrenska Academy at the University of Gothenburg, Sweden. Tel.: +46 31 342 1330; Fax: +46 31 82 84 58; E-mail: goran.larson@clinchem.gu.se.

Address Bruna Stråket 16, Sahlgrenska University Hospital, SE 413 45 Gothenburg, Sweden

During the revision of this manuscript, the *O*-glycosites of Protein delta homolog 1(Thr[256], Uniprot/KB accession P80370) and Protein YIPF3 (Thr[346], Uniprot/KB accession Q9GZM5) were independently identified by Steentoft *et al* Nat Methods, 2011, Oct 9. doi: 10.1038/nmeth.1731.

REFERENCES

1. Hoorn, E. J., Pisitkun, T., Zietse, R., Gross, P., Frokiaer, J., Wang, N. S., Gonzales, P. A., Star, R. A., and Knepper, M. A. (2005) Prospects for urinary proteomics: exosomes as a source of urinary biomarkers. *Nephrology* **10,** 283–290

2. Thongboonkerd, V., and Malasit, P. (2005) Renal and urinary proteomics: current applications and challenges. *Proteomics* **5,** 1033–1042

3. Heine, G., Raida, M., and Forssmann, W. G. (1997) Mapping of peptides and protein fragments in human urine using liquid chromatography-mass spectrometry. *J. Chromatogr. A* **776,** 117–124

4. Spahr, C. S., Davis, M. T., McGinley, M. D., Robinson, J. H., Bures, E. J., Beierle, J., Mort, J., Courchesne, P. L., Chen, K., Wahl, R. C., Yu, W., Luethy, R., and Patterson, S. D. (2001) Towards defining the urinary proteome using liquid chromatography-tandem mass spectrometry. I. Profiling an unfractionated tryptic digest. *Proteomics* **1,** 93–107

5. Pieper, R., Gatlin, C. L., McGrath, A. M., Makusky, A. J., Mondal, M., Seonarain, M., Field, E., Schatz, C. R., Estock, M. A., Ahmed, N., Anderson, N. G., and Steiner, S. (2004) Characterization of the human urinary proteome: a method for high-resolution display of urinary proteins on two-dimensional electrophoresis gels with a yield of nearly 1400 distinct protein spots. *Proteomics* **4,** 1159–1174

6. Pisitkun, T., Shen, R. F., and Knepper, M. A. (2004) Identification and proteomic profiling of exosomes in human urine. *Proc. Natl. Acad. Sci. U. S. A.* **101,** 13368–13373

7. Castagna, A., Cecconi, D., Sennels, L., Rappsilber, J., Guerrier, L., Fortis, F., Boschetti, E., Lomas, L., and Righetti, P. G. (2005) Exploring the hidden human urinary proteome via ligand library beads. *J. Proteome Res.* **4,** 1917–1930

8. Adachi, J., Kumar, C., Zhang, Y., Olsen, J. V., and Mann, M. (2006) The human urinary proteome contains more than 1500 proteins, including a large proportion of membrane proteins. *Genome Biol.* **7,** R80

9. Kentsis, A., Monigatti, F., Dorff, K., Campagne, F., Bachur, R., and Steen, H. (2009) Urine proteomics for profiling of human disease using high accuracy mass spectrometry. *Proteomics Clin. Appl.* **3,** 1052–1061

10. Nagaraj, N., and Mann, M. (2011) Quantitative analysis of the intra- and inter-individual variability of the normal urinary proteome. *J. Proteome Res.* **10,** 637–645

11. Spiro, R. G. (2002) Protein glycosylation: nature, distribution, enzymatic formation, and disease implications of glycopeptide bonds. *Glycobiology* **12,** 43R–56R

12. Varki, A. C. R., Esko, J. D., Freeze, H. H., Stanley, P., Bertozzi, C. R., Hart, G. W., and Etzler, M. E., editors. (2009) *Essentials of Glycobiology*, 2nd Ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY

13. Tian, E., and Ten Hagen, K. G. (2009) Recent insights into the biological roles of mucin-type O-glycosylation. *Glycoconj. J.* **26,** 325–334

14. Janik, M. E., Litynska, A., and Vereecken, P. (2010) Cell migration-the role of integrin glycosylation. *Biochim. Biophys. Acta* **1800,** 545–555

15. Tabak, L. A. (2010) The role of mucin-type O-glycans in eukaryotic development. *Semin. Cell Dev. Biol.* **21,** 616–621

16. Schjoldager, K. T., Vester-Christensen, M. B., Goth, C. K., Petersen, T. N., Brunak, S., Bennett, E. P., Levery, S. B., and Clausen, H. (2011) A systematic study of site-specific GalNAc-Type O-glycosylation modulating proprotein convertase processing. *J. Biol. Chem.*

17. Wang, L., Li, F., Sun, W., Wu, S., Wang, X., Zhang, L., Zheng, D., Wang, J., and Gao, Y. (2006) Concanavalin A-captured glycoproteins in healthy human urine. *Mol. Cell. Proteomics* **5,** 560–562

18. Sleat, D. E., Zheng, H., and Lobel, P. (2007) The human urine mannose 6-phosphate glycoproteome. *Biochim. Biophys. Acta* **1774,** 368–372

19. Moon, P. G., Hwang, H. H., Boo, Y. C., Kwon, J., Cho, J. Y., and Baek, M. C. (2008) Identification of rat urinary glycoproteome captured by three lectins using gel and LC-based proteomics. *Electrophoresis* **29,** 4324–4331

20. Yang, N., Feng, S., Shedden, K., Xie, X., Liu, Y., Rosser, C. J., Lubman, D. M., and Goodison, S. (2011) Urinary glycoprotein biomarker discovery for bladder cancer detection using LC/MS-MS and label-free quantification. *Clin. Cancer Res.* **17,** 3349–3359

21. Balog, C. I., Mayboroda, O. A., Wuhrer, M., Hokke, C. H., Deelder, A. M., and Hensbergen, P. J. (2010) Mass spectrometric identification of aberrantly glycosylated human apolipoprotein C-III peptides in urine from Schistosoma mansoni-infected individuals. *Mol. Cell. Proteomics* **9,** 667–681

22. Telford, J. E., Doherty, M. A., Tharmalingam, T., and Rudd, P. M. (2011) Discovering new clinical markers in the field of glycomics. *Biochem. Soc. Trans.* **39,** 327–330

23. Alonzi, D. S., Su, Y. H., and Butters, T. D. (2011) Urinary glycan markers for disease. *Biochem. Soc. Trans.* **39,** 393–398

24. Valmu, L., Alfthan, H., Hotakainen, K., Birken, S., and Stenman, U. H. (2006) Site-specific glycan analysis of human chorionic gonadotropin beta-subunit from malignancies and pregnancy by liquid chromatography–electrospray mass spectrometry. *Glycobiology* **16,** 1207–1218

25. Ramirez-Llanelis, R., Llop, E., Ventura, R., Segura, J., and Gutierrez-Gallego, R. (2008) Can glycans unveil the origin of glycoprotein hormones? - human chorionic gonadotrophin as an example. . *Mass Spectrom.* **43,** 936–948

26. Medzihradszky, K. F., Maltby, D. A., Hall, S. C., Settineri, C. A., and Burlingame, A. L. (1994) Characterization of protein N-glycosylation by reversed-phase microbore liquid chromatography/electrospray mass spectrometry, complementary mobile phases, and sequential exoglycosidase digestion. *J. Am. Soc. Mass Spectrom.* **5,** 350–358

27. Annesley, T. M. (2003) Ion suppression in mass spectrometry. *Clin. Chem.* **49,** 1041–1044

28. Peterman, S. M., and Mulholland, J. J. (2006) A Novel Approach for Identification and Characterization of Glycoproteins Using a Hybrid Linear Ion Trap/FT-ICR Mass Spectrometer. *J. Am. Soc. Mass Spectrom.* **17,** 168–179

29. Zhang, H., Li, X. J., Martin, D. B., and Aebersold, R. (2003) Identification and quantification of N-linked glycoproteins using hydrazide chemistry, stable isotope labeling and mass spectrometry. *Nat. Biotechnol.* **21,** 660–666

30. Bunkenborg, J., Pilch, B. J., Podtelejnikov, A. V., and Wisniewski, J. R. (2004) Screening for N-glycosylated proteins by liquid chromatography mass spectrometry. *Proteomics* **4,** 454–465

31. Hagglund, P., Bunkenborg, J., Elortza, F., Jensen, O. N., and Roepstorff, P. (2004) A new strategy for identification of N-glycosylated proteins and unambiguous assignment of their glycosylation sites using HILIC enrichment and partial deglycosylation. *J. Proteome Res.* **3,** 556–566

32. Palmisano, G., Lendal, S. E., Engholm-Keller, K., Leth-Larsen, R., Parker, B. L., and Larsen, M. R. (2010) Selective enrichment of sialic acid-containing glycopeptides using titanium dioxide chromatography with analysis by HILIC and mass spectrometry. *Nat. Protoc.* **5,** 1974–1982

33. Zielinska, D. F., Gnad, F., Wisniewski, J. R., and Mann, M. (2010) Precision mapping of an in vivo N-glycoproteome reveals rigid topological and sequence constraints. *Cell* **141,** 897–907

34. Kurogochi, M., Amano, M., Fumoto, M., Takimoto, A., Kondo, H., and Nishimura, S. (2007) Reverse glycoblotting allows rapid-enrichment glycoproteomics of biopharmaceuticals and disease-related biomarkers. *Angew Chem. Int. Ed. Engl.* **46,** 8808–8813

35. Kurogochi, M., Matsushista, T., Amano, M., Furukawa, J., Shinohara, Y., Aoshima, M., and Nishimura, S. (2010) Sialic acid-focused quantitative mouse serum glycoproteomics by multiple reaction monitoring assay. *Mol. Cell. Proteomics* **9,** 2354–2368

36. Jensen, P. H., Kolarich, D., and Packer, N. H. (2010) Mucin-type O-glycosylation–putting the pieces together. *FEBS J.* **277,** 81–94

37. Nilsson, J., Ruetschi, U., Halim, A., Hesse, C., Carlsohn, E., Brinkmalm, G., and Larson, G. (2009) Enrichment of glycopeptides for glycan structure and attachment site identification. *Nat. Methods* **6,** 809–811

38. Kelleher, N. L., Zubarev, R. A., Bush, K., Furie, B., Furie, B. C., McLafferty, F. W., and Walsh, C. T. (1999) Localization of labile posttranslational modifications by electron capture dissociation: the case of gamma-carboxyglutamic acid. *Anal. Chem.* **71,** 4250–4253

39. Mirgorodskaya, E., Roepstorff, P., and Zubarev, R. A. (1999) Localization of O-glycosylation sites in peptides by electron capture dissociation in a Fourier transform mass spectrometer. *Anal. Chem.* **71,** 4431–4436

40. Syka, J. E., Coon, J. J., Schroeder, M. J., Shabanowitz, J., and Hunt, D. F. (2004) Peptide and protein sequence analysis by electron transfer dissociation mass spectrometry. *Proc. Natl. Acad. Sci. U. S. A.* **101,** 9528–9533

41. Deguchi, K., Ito, H., Baba, T., Hirabayashi, A., Nakagawa, H., Fumoto, M., Hinou, H., and Nishimura, S. (2007) Structural analysis of O-glycopeptides employing negative- and positive-ion multi-stage mass spectra obtained by collision-induced and electron-capture dissociations in lin-

ear ion trap time-of-flight mass spectrometry. *Rapid Commun. Mass Spectrom.* **21,** 691–698

42. Perdivara, I., Petrovich, R., Allinquant, B., Deterding, L. J., Tomer, K. B., and Przybylski, M. (2009) Elucidation of O-glycosylation structures of the beta-amyloid precursor protein by liquid chromatography-mass spectrometry using electron transfer dissociation and collision induced dissociation. *J. Proteome Res.* **8,** 631–642

43. Sihlbom, C., van Dijk Hard, I., Lidell, M. E., Noll, T., Hansson, G. C., and Backstrom, M. (2009) Localization of O-glycans in MUC1 glycoproteins using electron-capture dissociation fragmentation mass spectrometry. *Glycobiology* **19,** 375–381

44. Christiansen, M. N., Kolarich, D., Nevalainen, H., Packer, N. H., and Jensen, P. H. (2010) Challenges of determining O-glycopeptide heterogeneity: a fungal glucanase model system. *Anal. Chem.* **82,** 3500–3509

45. Takahashi, K., Wall, S. B., Suzuki, H., Smith, A. D. T., Hall, S., Poulsen, K., Kilian, M., Mobley, J. A., Julian, B. A., Mestecky, J., Novak, J., and Renfrow, M. B. (2010) Clustered O-glycans of IgA1: defining macro- and microheterogeneity by use of electron capture/transfer dissociation. *Mol. Cell. Proteomics* **9,** 2545–2557

46. Darula, Z., and Medzihradszky, K. F. (2009) Affinity enrichment and characterization of mucin core-1 type glycopeptides from bovine serum. *Mol. Cell. Proteomics* **8,** 2515–2526

47. Forsman, A., Ruetschi, U., Ekholm, J., and Rymo, L. (2008) Identification of intracellular proteins associated with the EBV-encoded nuclear antigen 5 using an efficient TAP procedure and FT-ICR mass spectrometry. *J. Proteome Res.* **7,** 2309–2319

48. Strohalm, M., Hassman, M., Kosata, B., and Kodicek, M. (2008) mMass data miner: an open source alternative for mass spectrometric data analysis. *Rapid Commun. Mass Spectrom.* **22,** 905–908

49. Jung, E., Veuthey, A. L., Gasteiger, E., and Bairoch, A. (2001) Annotation of glycoproteins in the SWISS-PROT database. *Proteomics* **1,** 262–268

50. Domon, B., and Costello, C. E. (1988) A systematic nomenclature for carbohydrate fragmentations in FAB-MS/MS spectra of glycoconjugates. *Glycoconjugate J.* **5,** 397–409

51. Morris, H. R., Chalabi, S., Panico, M., Sutton-Smith, M., Clark, G. F., Goldberg, D., and Dell, A. (2007) Glycoproteomics: Past, present and future. *Int. J. Mass Spectrom.* **259,** 16–31

52. Hudgins, W. R., Hampton, B., Burgess, W. H., and Perdue, J. F. (1992) The identification of O-glycosylated precursors of insulin-like growth factor II. *J. Biol. Chem.* **267,** 8153–8160

53. Mormann, M., Paulsen, H., and Peter-Katalinic, J. (2005) Electron capture dissociation of O-glycosylated peptides: radical site-induced fragmentation of glycosidic bonds. *Eur. J. Mass Spectrom.* **11,** 497–511

54. D'Amico, G., and Bazzi, C. (2003) Pathophysiology of proteinuria. *Kidney Int.* **63,** 809–825

55. Thongboonkerd, V., Chutipongtanate, S., and Kanlaya, R. (2006) Systematic evaluation of sample preparation methods for gel-based human urinary proteomics: quantity, quality, and variability. *J. Proteome Res.* 5, 183–191

56. Huddleston, M. J., Bean, M. F., and Carr, S. A. (1993) Collisional fragmentation of glycopeptides by electrospray ionization LC/MS and LC/MS/MS: methods for selective detection of glycopeptides in protein digests. *Anal. Chem.* **65,** 877–884

57. Wuhrer, M., Koeleman, C. A., and Deelder, A. M. (2009) Hexose rearrangements upon fragmentation of N-glycopeptides and reductively aminated N-glycans. *Anal. Chem.* **81,** 4422–4432

58. Wuhrer, M., Deelder, A. M., and van der Burgt, Y. E. (2011) Mass spectrometric glycan rearrangements. *Mass Spectrom. Rev.* **30,** 664–680

59. Halim, A., Brinkmalm, G., Ruetschi, U., Westman-Brinkmalm, A., Portelius, E., Zetterberg, H., Blennow, K., Larson, G., and Nilsson, J. (2011) Site-specific characterization of threonine, serine, and tyrosine glycosylations of amyloid precursor protein/amyloid {beta}-peptides in human cerebrospinal fluid. *Proc. Natl. Acad. Sci. U. S. A.* **108,** 11848–11853

60. Wada, Y., Azadi, P., Costello, C. E., Dell, A., Dwek, R. A., Geyer, H., Geyer, R., Kakehi, K., Karlsson, N. G., Kato, K., Kawasaki, N., Khoo, K. H., Kim, S., Kondo, A., Lattova, E., Mechref, Y., Miyoshi, E., Nakamura, K., Narimatsu, H., Novotny, M. V., Packer, N. H., Perreault, H., Peter-Katalinic, J., Pohlentz, G., Reinhold, V. N., Rudd, P. M., Suzuki, A., and Taniguchi, N. (2007) Comparison of the methods for profiling glycoprotein glycans–HUPO Human Disease Glycomics/Proteome Initiative multi-institutional study. *Glycobiology* **17,** 411–422

61. Easton, R. L., Patankar, M. S., Clark, G. F., Morris, H. R., and Dell, A. (2000) Pregnancy-associated changes in the glycosylation of tamm-horsfall glycoprotein. Expression of sialyl Lewis(x) sequences on core 2 type O-glycans derived from uromodulin. *J. Biol. Chem.* **275,** 21928–21938

62. Watson, J. A., Burling, K., Fitzpatrick, P., Kay, E., Kelly, J., Fitzpatrick, J. M., Dervan, P. A., and McCann, A. (2009) Urinary insulin-like growth factor 2 identifies the presence of urothelial carcinoma of the bladder. *BJU Int.* **103,** 694–697

63. Kobayashi, H., Terao, T., and Kawashima, Y. (1992) Serum sialyl Tn as an independent predictor of poor prognosis in patients with epithelial ovarian cancer. *J. Clin. Oncol.* **10,** 95–101

64. David, L., Nesland, J. M., Clausen, H., Carneiro, F., and Sobrinho-Simoes, M. (1992) Simple mucin-type carbohydrate antigens (Tn, sialosyl-Tn and T) in gastric mucosa, carcinomas and metastases. *APMIS Suppl.* **27,** 162–172

65. Itzkowitz, S. H., Bloom, E. J., Kokal, W. A., Modin, G., Hakomori, S., and Kim, Y. S. (1990) Sialosyl-Tn. A novel mucin antigen associated with prognosis in colorectal cancer patients. *Cancer* **66,** 1960–1966

66. Kim, G. E., Bae, H. I., Park, H. U., Kuan, S. F., Crawley, S. C., Ho, J. J., and Kim, Y. S. (2002) Aberrant expression of MUC5AC and MUC6 gastric mucins and sialyl Tn antigen in intraepithelial neoplasms of the pancreas. *Gastroenterology* **123,** 1052–1060

67. Springer, G. F. (1989) Tn epitope (N-acetyl-D-galactosamine alpha-O-serine/threonine) density in primary breast carcinoma: a functional predictor of aggressiveness. *Mol. Immunol.* **26,** 1–5

68. Muller, S., Alving, K., Peter-Katalinic, J., Zachara, N., Gooley, A. A., and Hanisch, F. G. (1999) High density O-glycosylation on tandem repeat peptide from secretory MUC1 of T47D breast cancer cells. *J. Biol. Chem.* **274,** 18165–18172